

Numerical algebraic geometry for model selection and its application to the life sciences

Elizabeth Gross¹ Brent Davis², Kenneth L. Ho³, Daniel J. Bates², Heather A. Harrington⁴

¹Department of Mathematics, San José State University,

²Department of Mathematics, Colorado State University,

³Department of Mathematics, Stanford University,

⁴Mathematical Institute, University of Oxford

April 4, 2016

Abstract

Researchers working with mathematical models are often confronted by the related problems of parameter estimation, model validation, and model selection. These are all optimization problems, well-known to be challenging due to non-linearity, non-convexity and multiple local optima. Furthermore, the challenges are compounded when only partial data is available. Here, we consider polynomial models (e.g., mass-action chemical reaction networks at steady state) and describe a framework for their analysis based on optimization using numerical algebraic geometry. Specifically, we use probability-one polynomial homotopy continuation methods to compute all critical points of the objective function, then filter to recover the global optima. Our approach exploits the geometric structures relating models and data, and we demonstrate its utility on examples from cell signaling, synthetic biology, and epidemiology.

1 Introduction

Across the physical, biological, and social sciences, mathematical models are formulated and studied to better understand real-world phenomena. Often, multiple models are developed to explore alternate hypotheses. It then becomes necessary to choose between different models, for example, based on their fit with noisy experimental data. This is the problem of model selection, a fundamental scientific problem with practical implications (1–3).

The standard approach to model selection is to first estimate all model parameters and hidden variables from the data, then select the model with the smallest best-fit error, up to some penalty on model complexity (4, 5). Thus, at its core, model selection is intimately tied to parameter estimation, which, for a given model $f(a, x) = 0$ in the parameters a and variables x with measurable “outputs” $y = g(x)$, say, can be written as the following optimization problem:

$$\min_{a, x, y} \|y - y'\|^2 \quad \text{s.t.} \quad \begin{cases} f(a, x) = 0 \\ y = g(x) \end{cases} \quad (1)$$

where y' denotes the observed data, i.e. measured outputs. Unless f and g are convex, solving (1) is a nonconvex problem, which can be challenging as standard local solvers run the risk of getting trapped in local minima (especially in high dimensions). This can be mitigated somewhat with techniques such as simulated annealing (6, 7) or convex relaxation which has been successful for model invalidation (8–10), but there is generally no guarantee that a global minimum will be found.

When f and g are polynomial, however, problem (1) can be solved globally by finding all roots of an associated polynomial system. In this case, ideas from computational algebra and algebraic geometry can be effective; see, e.g., (11–14) for applications of Gröbner bases in systems biology and (15) for applications of algebraic geometry to statistical inference. Such symbolic methods tend to be computationally expensive,

which limits their use in practice. Thus, although they provide a solution in principle, new algorithms and techniques are yet desired.

In this paper, we aim to fill this gap by proposing a framework for global parameter estimation for polynomial deterministic models using numerical algebraic geometry (NAG), a suite of tools for numerically approximating the solution sets of multivariate polynomial systems via adaptive multi-precision, probability-one polynomial homotopy continuation (16, 17). Our approach scales well in dimension relative to classical symbolic methods (18), and, while it comes with a higher computational cost than standard local optimization, it has a probability-one guarantee to recover the global optima, solving problem (1) in the strong sense. This allows us to reason rigorously about model fit and to address the related problems of model selection and parameter estimation from a maximum-likelihood perspective.

We demonstrate our techniques on examples from biology, where polynomial models often arise as the steady-state descriptions of mass-action chemical reaction networks. Although some limitations remain, we believe that this work achieves its primary purpose of introducing NAG as a valuable complement to existing tools for model evaluation and analysis. Additionally, this paper highlights specific challenges that arise when using polynomial methods for model inference, such as dealing with positivity constraints and non-isolated solutions, and provides guidance for tackling these challenges.

The remainder of the paper is organized as follows. In the next section, we state precisely the problems with which we are concerned: model validation, model selection, and parameter and hidden variable estimation. We then present NAG algorithms for solving each problem. Finally, we showcase our approach on a few examples including cell death activation, synthetic bio-circuits, HIV progression, and protein modification.

2 Problem Statement

Consider a model whose dynamics are described by the system of polynomial differential equations

$$\dot{x} = f(a, x) \quad (2)$$

where $a = (a_1, \dots, a_k)$ are parameters (e.g. rate constants in a deterministic mechanistic model, such as a chemical reaction network with mass-action kinetics), $x = (x_1, \dots, x_n)$ are variables, and $f = (f_1, \dots, f_r)$ are polynomials in x and a with measurable outputs $y = g(x)$ where $y = (y_1, \dots, y_m)$, $m \leq n$, and $g = (g_1, \dots, g_m)$ are polynomials in x . While the parameters a_1, \dots, a_k are treated as fixed variables in our exposition, we separate them from x_1, \dots, x_n to respect how such variables are treated differently in experimental and computational settings.

In algebraic geometry a *variety* is a solution set of a system of polynomial equations; we use this terminology for our next two definitions. The *real model variety* is the solution set of the system

$$f = 0 \quad (3)$$

$$y - g(x) = 0, \quad \text{that is,} \quad (4)$$

$$(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} := \{(a, x, y) \in \mathbb{R}^{k+n+m} : f(a, x) = 0, y - g(x) = 0\}$$

corresponding to the steady states of the model. Now, consider, for simplicity, the case of a single data point $\hat{y} = (\hat{y}_1, \dots, \hat{y}_m)$. (See SI Appendix for multiple data points.) The *real data variety* is then the affine linear space

$$(\mathcal{V}_{\mathcal{D}})_{\mathbb{R}} := \{(a, x, y) \in \mathbb{R}^{k+n+m} : y_i = \hat{y}_i, \forall i = 1, \dots, m\},$$

with $\dim(\mathcal{V}_{\mathcal{D}})_{\mathbb{R}} = k + n$. We consider the case when the data includes some extrinsic (measurement) noise; we assume the errors $\{\epsilon_1, \dots, \epsilon_m\}$ on the observed data variables are uncorrelated random variables and each error ϵ_i is normally distributed with known variance σ_i (which can be obtained by instrument calibration).

Using this geometric framework, the problems of (1) model validation, (2) model selection, and (3) parameter estimation, can be described precisely in terms of the real algebraic varieties $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ and $(\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$.

2.1 Problem 1: Model Validation

For model validation, we want to determine whether a deterministic polynomial model \mathcal{M} is compatible with the data according to a given significance level α . This is akin to asking whether the model is a “good fit” for the data. A natural goodness-of-fit statistic is:

$$d^2 := \min \sum_{i=1}^m \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2} \text{ subject to } (a, x, y) \in (\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}. \quad (5)$$

When the variances differ, d^2 is the minimum squared weighted Euclidean distance between $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ and $(\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$. When all variances are equal to one, the value of (5) is just the minimum squared distance. For the remainder of the text, we will assume the latter, knowing that we can simply rescale.

Optimization problem 5 can be derived directly from the log-likelihood function

$$\log L(a, x, y) = \sum_{i=1}^m \left(\frac{1}{2} \log 2\pi\sigma_i^2 - \frac{(\hat{y}_i - y_i)^2}{2\sigma_i^2} \right)$$

as demonstrated in Section 2.1 of the SI. This provides a bridge to standard statistical model selection tools such as AIC (19) and BIC (20).

If the data are generated from the model \mathcal{M} with normally distributed extrinsic noise, the distribution function of d^2 is dominated by that of the chi-squared distribution with m degrees of freedom, χ_m^2 , where m is the number of measurable outputs. Thus, if p_α is the upper α -percentile for χ_m^2 , then $\Pr(d^2 \geq p_\alpha) \leq \Pr(U \geq p_\alpha) = \alpha$ where $U \sim \chi_m^2$. We reject the model \mathcal{M} as *incompatible* if $d^2 > p_\alpha$; otherwise we say that the model \mathcal{M} is *compatible*.

If the real model and data varieties intersect, that is, $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}} \neq \emptyset$, then $d^2 = 0$, and we also say that the model is compatible with the data. If there are restrictions on (a, x, y) , for example if all parameters and variables are required to be non-negative, then finding d^2 becomes a *constrained optimization* problem (see SI Appendix).

2.2 Problem 2: Model Selection

For model selection, we are given a set of models $\{\mathcal{M}_1, \dots, \mathcal{M}_s\}$ and want to determine the model of best fit. In this setting, we use the statistic d^2 to make a selection, either by choosing the model with the smallest goodness-of-fit statistic or by using d^2 in conjunction with a complexity penalty, similar to the Bayesian or Akaike information criteria (2).

If the statistic d^2 evaluates to zero for all (or even multiple) models, then we are unable to make a selection between the models. This can be remedied by designing experiments that yield more informative data. For example, measuring more variables can reduce the dimension of $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$; the most informative situation is when $(\mathcal{V}_{\mathcal{M}_i})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}} = \emptyset$ for all models. This indeterminacy can also be resolved by taking multiple measurements and minimizing the joint squared distance (see SI Appendix).

2.3 Problem 3: Parameter Estimation

Parameter estimation can be achieved by finding the point $(a^*, x^*, y^*) \in (\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ that minimizes the value of (5). The point (a^*, x^*, y^*) is the maximum likelihood estimate under the given noise assumptions. The parameter estimate is then a^* ; the estimate of the hidden variables, x^* ; and the estimate of the “de-noised” outputs, y^* . Of course, if the data and model varieties intersect, there will be one or more (possibly infinite) choices for (a^*, x^*, y^*) . Otherwise, it is a matter of solving a polynomial system that yields the point(s) on $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ nearest $(\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$. This is described in more detail in the next section.

3 Geometry

In each of the problems stated above, we seek to minimize the distance between $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ and $(\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$ or to find the intersection of these two sets. Standard methods for solving non-linear optimization problems are

local in nature, i.e., only guaranteed to converge to a local minimum which may or may not be the global minimum. However, using NAG, we find *all* local extrema over \mathbb{C} , necessarily including the global minimum. Due to this global benefit, NAG has been used before in statistical inference in the field of algebraic statistics (21, 22).

Let $V_{\mathcal{M}} \subseteq \mathbb{C}^{k+n+m}$ be the (complex) Zariski closure of $(V_{\mathcal{M}})_{\mathbb{R}}$ and $V_{\mathcal{D}} \subseteq \mathbb{C}^{k+n+m}$ be the (complex) Zariski closure of $(V_{\mathcal{D}})_{\mathbb{R}}$. We will refer to $V_{\mathcal{M}}$ and $V_{\mathcal{D}}$ as the *model variety* and *data variety*, respectively.

The problem of determining the intersection of $V_{\mathcal{M}} \cap V_{\mathcal{D}}$ is simply a matter of solving the polynomial system obtained by taking the union of the polynomials defining $V_{\mathcal{M}}$ and the polynomials defining $V_{\mathcal{D}}$. This is handled directly by NAG. If the intersection is nonempty and positive-dimensional (complex curves, surfaces, etc), real points can be found using the polynomial homotopy method described in (23), a method based on symbolic algorithms in (24), and, more classically, on the decision method in (25).

The problem of finding the points on the two varieties nearest one another can also be stated in terms of a polynomial system, on which we then call NAG solvers to find solutions. A well-known necessary condition for local extrema is given by Lagrange multipliers. In the main text we assume that $r + m = \text{codim } \mathcal{V}_{\mathcal{M}}$; however, when this is not the case, the number of equations can be reduced (see SI Appendix).

Proposition 1 (Equations given by Lagrange multipliers). *Let $c = \text{codim } V_{\mathcal{M}}$ and let $f' = (f'_1, \dots, f'_c) = 0$ on a Zariski open set of $V_{\mathcal{M}}$. If $(a, x, y) \in (V_{\mathcal{M}})_{\mathbb{R}}$ is a local minimum of*

$$\sum_{i=1}^m (y_i - \hat{y}_i)^2, \quad (6)$$

then there exists $\lambda := (\lambda_1, \dots, \lambda_c)$, such that (a, x, y, λ) is a solution to the system

$$f' = 0, \quad (7)$$

$$y - g(x) = 0 \quad (8)$$

$$\lambda_1 \nabla f'_1 + \dots + \lambda_c \nabla f'_c + \begin{bmatrix} 0 \\ y - \hat{y} \end{bmatrix} = 0. \quad (9)$$

Solving this system with NAG will provide us with all local extrema of (6) over $\mathcal{V}_{\mathcal{M}}$, from which we may easily select the pair of nearest points. The geometry of zero sets of systems such as (7)-(9) are described in (26).

3.1 Numerical Algebraic Geometry

Given a polynomial system F consisting of r polynomials and N variables, NAG packages, such as **Bertini** (17), **PHCpack** (27), **HOM4PS-2.0** (28), use polynomial homotopy continuation to provide *probability-one* numerical methods for finding approximations of all isolated complex solutions of $F = 0$ (points) as well as *witness points* on all positive-dimensional irreducible components of the solution set of $F = 0$. These methods are probability-one in that the computations include some randomization, and this randomization will yield theoretically correct results so long as the random numbers are not chosen from some measure zero set in the parameter spaces of potential choices (16, 29).

If $x \in \mathbb{R}^N$ is a *real* solution of $F = 0$, it is either isolated among the complex solutions or it lies on a positive-dimensional complex irreducible component. In the former case, the methods of NAG will find x and recognize it as real. In the latter case, x can be difficult to uncover. However, for our purposes, we usually only need to verify the existence of a real solution. In this case, we can find witness points on all positive dimensional components and then use the procedure in Section 2.1 of (23) to verify the existence of real points.

3.2 Algorithms

We present three related algorithms to address model validation, model selection, and parameter estimation (see Fig. 1).

The aim of the first algorithm, Algorithm 1, is to find the pair of points that minimize the distance between $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ and $(\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$. If $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}} = \emptyset$, then this is obtained by solving (7)-(9); otherwise, additional techniques are required.

We note also simply that these techniques are indeed probability-one algorithms. The computations will necessarily be carried out in finite time and the steps proceed linearly (no loops), so the methods will necessarily finish.

3.2.1 Algorithm 1: Model validation

Input	Model \mathcal{M} , data $\mathcal{D} = \{\hat{y}\}$, tolerance α
Output	<i>yes</i> or <i>no</i>
1	If $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}} = \emptyset$, goto Step 3.
2	If $\dim(\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}) \geq 0$ and $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}} \neq \emptyset$, return <i>yes</i> , else, goto Step 3.
3	Find a pair $(z_1, z_2) \in (\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \times (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$ that minimizes (6) (using NAG software such as Bertini or PHCpack).
4	If $\ z_1 - z_2\ ^2 < p_{\alpha}$, return <i>yes</i> ; else <i>no</i> .

The computation of the intersection $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ in Steps 1 and 2 can be determined in several ways. The simplest way is by considering dimensions: if $\dim(\mathcal{V}_{\mathcal{M}}) + \dim(\mathcal{V}_{\mathcal{D}})$ exceeds the ambient dimension, they will almost surely intersect. If the ambient dimension is larger than the sum of the variety dimensions, they typically do not intersect. To compute the intersection (or check to see if it is nonempty), one could substitute $\hat{y} - g(x)$ for (8) in the system (7)-(8).

In Step 2, if $\dim(\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}) = 0$, then the intersection of the two varieties consists of finitely many points; the condition $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}} \neq \emptyset$ indicates that at least one of the points is real, which is straightforward to determine. If $\dim(\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}) > 0$, to check if $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}} \neq \emptyset$, one needs more sophisticated methods, such as those in (23).

To find the pair (z_1, z_2) in Step 3, one may solve the polynomial system (7)-(9). If there is a positive-dimensional set of (complex) extremal points, then the procedures in (23) could be used to determine if the set contains a real point.

If there are constraints on the variables or parameters, for example, if we seek to minimize (6) over the non-negative part of $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$, then the algorithm is updated as follows: if the $\dim(\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}) = 0$ or $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}} = \emptyset$ then the Karush-Kuhn-Tucker equations as described in Section 3.1 of the SI Appendix should be used, while if $\dim(\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}) > 0$, or if there is a positive-dimensional set of extremal points at Step 3, then the algorithm should return *possibly*. There are methods (30–32) for finding real points, curves, and surfaces within complex components of dimension 2 or less, but little is known about higher dimensions.

3.2.2 Algorithm 2: Model selection

In this case, there are several competing models, each with its own polynomial system. The algorithm proceeds much as in Algorithm 1, but iterated for the several models under consideration. If a threshold α is provided, one should first reject models that do not adequately support the data ($d^2 \geq p_{\alpha}$), then choose the model yielding the minimum value of d^2 (up to some complexity penalty). Various conclusions may be drawn, e.g., *no model adequately fits the data* or *three models adequately fit the data and model \mathcal{M}_1 provides the best fit*.

3.2.3 Algorithm 3: Parameter estimation

Again, this algorithm is similar to the first. The input consists of only one model \mathcal{M} and data \mathcal{D} . It is assumed that there are unknown parameters and the goal is to find values of these parameters producing the best fit between \mathcal{M} and \mathcal{D} . The output of Steps 2 and 3 need to be adjusted appropriately. The output of Step 4 is simply z_2 since there is no α to be used for rejection. The method also simultaneously estimates hidden/unknown variables and “de-noised” outputs.

3.2.4 Simple Example

To illustrate Algorithm 1, consider a simple model with three variables x, y, z and three parameters a, b, c satisfying

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

Let $\alpha = 0.1$ and assume, for this example, that the variances on the errors are $\sigma_i^2 = 0.1$ for $i = 1, 2, 3$. The model variety $\mathcal{V}_{\mathcal{M}}$ is simply an ellipsoid (Fig. 2A) where a, b, c describe the principal axes. Suppose we know that $a, b, c = 1$. For the case when the outputs are x and y and $\hat{x} = 0, \hat{y} = 0$ and $\alpha = 0.1$, Step 2 indicates that the data *does* fit the model, i.e. the model is compatible with the data. In this case, there are two real points in the 0-dimensional intersection $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ (see Fig. 2B). For the same set-up, but with data $\hat{x} = 0$, Step 3 indicates that this data *possibly* fits this model. Since there is a positive-dimensional intersection (Fig. 2C), it is possibly compatible (depending on constraints imposed by the user). For the same model and α , but different data $\hat{x} = 1.7, \hat{y} = 0$, Step 4 yields points $(1, 0, 0)$ and $(1.7, 0, 0)$, so $\|z_1 - z_2\|^2 > 0.4605 = p_\alpha$, and the model is rejected (Fig. 2D). However when the data are $\hat{x} = 1.01, \hat{y} = 0$, Algorithm 1 indicates model compatibility (Fig. 2E). Previous algebraic methods that required Gröbner basis calculations would result in a zero set and thus those approaches are not useful here.

4 Results

We demonstrate our methods on problems in biomedicine: cell death activation, synthetic biology, epidemiology and multisite phosphorylation. Each of the forthcoming applications can be written as a mass-action chemical reaction network, which has the form $\dot{x} = f(a, x)$, studied at steady state: $f(a, x) = 0$ as in the problem statement. Throughout these real-world examples, we emphasize the pivotal computations in these methods, such as determining the dimension of the intersection $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ and finding points in the intersection (see Fig. 1B). Our first two examples, cell death signaling and genetic toggle bio-circuits, demonstrate how to handle positive dimensional intersections using two different approaches, while the remaining two examples, HIV and MAP kinase signaling, highlight analysis of zero-dimensional and empty intersections. In the following examples, we are interested in results that can be interpreted biologically, therefore we restrict our analysis to non-negative real solutions.

4.1 Cell death activation

We demonstrate model compatibility (Algorithm 1) on an example from receptor-mediated programmed cell death, which is initiated by the activation of death receptors upon the detection of extracellular death ligands (33–36). We consider, in particular, the “cluster” model of (12), which was inspired by crystallographic data (37) and describes the recruitment of receptors by ligands into local self-activating clusters capable of bistability.

The cluster model is a system of 3 degree-4 polynomials in the form of (2) in 3 variables (representing various receptor states) and 6 rate parameters, supplemented by ligand and receptor conservations (see SI Appendix). We assume that we can measure the total ligand and receptor concentrations, which may be considered experimental inputs, as well as the concentration of active receptors. We do not assume access to the concentrations of other individual receptor states nor to any of the rate parameters.

A steady-state data point was simulated from the model with all parameters and initial concentrations drawn independently and identically from the log-normal distribution $\ln \mathcal{N}(0, 4)$, then combined and corrupted with i.i.d. noise from $\mathcal{N}(0, 0.1)$ to obtain \hat{y} . The real model and real data varieties intersect in the positive orthant with a distance zero and hence the model is indeed compatible with the data.

4.2 Synthetic biology and experimental design

We demonstrate an example from synthetic biology with excess intersection ($\dim(\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}) > 0$). A goal in synthetic biology is to design or modify existing biological systems with new features according to specific design criteria. Reverse engineering of biological systems often includes modules (such as feedback loops) and how these are interconnected are described by different circuits (models). Understanding differences between

bio-circuit implementations is crucial; therefore we compare three bistable bio-circuits models analyzed in (38): monomer-dimer toggle circuit (\mathcal{M}_1), dimer-dimer toggle circuit (\mathcal{M}_2), and single operator gene circuit (\mathcal{M}_3), which were initially presented in (39–41). The model variables include genes (X_i) and proteins (P_i) where $i = 1$ for \mathcal{M}_3 or $i = 1, 2$ for $\mathcal{M}_1, \mathcal{M}_2$, as well as these species complexes (e.g., $P_i P_i, X_j P_i P_i$). These variables interact following mass-action kinetics and form systems of polynomial differential equations where $\mathcal{M}_1, \mathcal{M}_2$, and \mathcal{M}_3 have 7, 8 and 6 model variables, respectively, and 10, 12, and 9 kinetic parameters, respectively. The models can be reduced (given in SI Appendix) by assuming that the total amount of gene 1 ($X_{1_{tot}}$) and gene 2 ($X_{2_{tot}}$) is conserved.

Suppose that the total amounts $X_{1_{tot}}$ and $X_{2_{tot}}$ and specific protein synthesis and degradation parameters $k_{bas_1}, k_{bas_2}, k_{deg_1}$, and k_{deg_2} are known. Since protein concentrations are often measurable, we assume that our data are P_1, P_2 , and their complexes $P_1 P_1$, and $P_2 P_2$. The aim is to select the best model $\mathcal{M}_1, \mathcal{M}_2$, and \mathcal{M}_3 given the data. We simulate steady-state data from the dimer-dimer toggle model (\mathcal{M}_2) and add Gaussian noise from $\mathcal{N}(0, 0.1)$. We find that all three have positive dimensional intersections, where the dimension of the intersections are:

$$\begin{aligned}\dim((\mathcal{V}_{\mathcal{M}_1})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}) &= 3, \\ \dim((\mathcal{V}_{\mathcal{M}_2})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}) &= 4, \\ \dim((\mathcal{V}_{\mathcal{M}_3})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}) &= 3.\end{aligned}$$

Clearly, all three models are compatible with the data, thus one can only select a “best fit” model using data-independent measures, e.g. number of parameters, dimension, etc.

In fact, the dimensions of the model and data varieties can help us design more informative experiments for model selection. For example, the variety associated to the monomer-dimer toggle model lives in a 15 dimensional ambient space, $\mathcal{V}_{\mathcal{M}_1} \subseteq \mathbb{C}^{15}$, and has dimension 6, while the data variety $\mathcal{V}_{\mathcal{D}} \subseteq \mathbb{C}^{15}$ has dimension 12; thus by the Dimension Theorem, we can determine that $\dim(\mathcal{V}_{\mathcal{M}_1} \cap \mathcal{V}_{\mathcal{D}}) \geq 6 + 12 - 15 = 3$. This dimension calculation provides guidance towards the number of minimal additional variable and parameter values that must be measured to ensure $\mathcal{V}_{\mathcal{M}_1} \cap \mathcal{V}_{\mathcal{D}} = \emptyset$, i.e., at least four more variables and parameter values must be known.

Suppose we can experimentally measure four forward biochemical reaction rate constants (e.g., k_{cF}, k_{kF}, k_{nF} , and k_{kR}), then $\mathcal{V}_{\mathcal{M}_1}$ is cut down by four dimensions and does not intersect $\mathcal{V}_{\mathcal{D}}$. We get similar results for the model varieties associated to \mathcal{M}_2 and \mathcal{M}_3 provided we measure rate constants specific to these models (see SI Appendix). Now that all the intersections are empty, we run Algorithm 2 and find that the sum of squares (Eq. (5)) for each model are as follows: $d_1^2 = 0.2262$, $d_2^2 = 7.34 \times 10^{-7}$, and $d_3^2 = 0.3040$. Therefore, we select the \mathcal{M}_2 model, which is indeed the true model.

4.3 HIV Progression

We demonstrate parameter estimation (Algorithm 3) on an example coming from epidemiology. We use a model that includes long-term HIV dynamics from initial viremia, latency, and virus increase (42), based on (43). In the model (see SI Appendix), the HIV virus inhibits the CD4+T cell population while promoting macrophage proliferation, which in turn houses the replicating virus. As macrophages proliferate, the virus reservoir increases, so the model offers a description of HIV patient progression to AIDS. Model variables x are uninfected CD4+T cells (T), infected CD4+ T cells (T_i), uninfected macrophages (M), infected macrophages (M_i), and HIV virus population (V).

Hernandez-Vargas et al. show that the model can have two real equilibria, one of which is stable, representing patients that are “long-term non-progressors” (42). The parameters a are $(s_1, s_2, k_1, \dots, k_6, \delta_1, \dots, \delta_5)$, where s_i represents synthesis of T cells and macrophages, k are rate constants describing interactions between variables x , and δ_i represents natural death. For this example, $y = x$. We estimated the natural death of the virus, parameter δ_5 , using the long-term non-progressors steady-state value (Table 3 of (42)) and adding noise to each variable $\sim \mathcal{N}(0, 1)$. By Algorithm 3, the data variety and model variety do not intersect. We find the closest point and estimate $\delta_5^* = 2.99876$ (true value of $\delta_5 = 3$).

4.4 Multisite phosphorylation with experimental data

We examine phosphorylation mechanisms of cellular signaling with experimental data, and demonstrate model selection (Algorithm 2) and parameter estimation (Algorithm 3). We focus on phosphorylation, a key cellular regulatory mechanism that has been the subject of extensive study, both experimentally and theoretically ((44) and references therein). An area of interest is the mechanism by which a kinase phosphorylates a two-site substrate, either distributively, where the kinase can add at most one phosphate before dissociating, or processively, where it can add both phosphates in sequence. The MAPK/ERK pathway is a well-known system for studying phosphorylation, whereby MEK (kinase) phosphorylates ERK (its substrate). Aoki et al. (45) showed experimental evidence, while working with polynomial models, that the mammalian MAPK/ERK pathway acts distributively *in vitro* but processively *in vivo*.

We compare these distributive and processive models against the *in vivo* data reported in the same study. The distributive model consists of 12 molecular species and 17 mass-action reactions, while the processive model has 14 species and 18 reactions (species correspond to variables, each reaction corresponds to a parameter). The data take the form of 36 concentration measurements of three aggregate phosphoforms over a range of 12 EGF stimulation levels. All model parameters are calibrated using *in vitro* estimates by (45), except the parameter k_1 representing EGF loading, which we estimate for both models (see SI Appendix).

Next we perform model selection by running Algorithm 2 on each data point individually and select independently for each run the preferred model (more details in SI Appendix). Under low EGF stimulation, the best model estimates are nearly identical with a slight preference for distributive. At high EGF stimulation, the models are identical with no preference for one model over the other. These results can be justified by noting that the main distinction between distributivity and processivity is nonlinear switching behavior (i.e., a sigmoidal response curve), and this only occurs at intermediate stimulations. However, at medium EGF stimulation (see Fig 3), there is a preference for the processive model, which supports the findings in (45).

5 Conclusion

The problem of determining whether given real-world data fits one or more given mathematical models is challenging. When a model is defined by algebraic (polynomial) functions, the methods of NAG may be employed to study the geometry underlying the model and data. In particular, these methods are useful for model variables observed at steady-state. We demonstrated this numerical and geometric framework for comparing models with experimental dose-response data in MAPK/ERK pathway and highlighted that the intermediate EGF doses were the most informative for model selection, complementing another finding that model selection results can be very sensitive to experimental parameters (46).

Despite the difficulties associated with positive dimensional components, and limitations in analysis, we can reproduce compatible models, and furthermore can predict additional information, such as measurements of parameters, that are necessary for selecting models. Our geometric investigations of positive dimensional components may perhaps relate to algebraic analyses for biochemical models, such as model identifiability or matroids for experimental design (14, 47, 48).

There are further directions to be considered in this vein, aside from making the existing computational methods more efficient. First, there would be great value in developing strictly *real* geometric methods for solving polynomial systems such as those that appear in this article. Some such techniques exist, but only in very special circumstances. Second, there would be much value in developing effective numerical methods for treating inequalities. It should be noted that the methods described in (49) and the references therein will incorporate such constraints, though the cost of such computations restricts their use to relatively low dimensions. Finally, there is likely much to be gained from considering the geometry underlying models not defined by algebraic functions. Algebraic geometry provides very clean, well-understood structures, paving the way for numerical methods. Differential geometry or topology could lead to similarly useful techniques for model selection and parameter estimation.

6 Materials and Methods

6.1 Numerical algebraic geometry

General references for NAG include (16, 29), with the latter doubling as a user manual for the software package Bertini. For computations, we used Bertini 1.4 and Macaulay2 version 1.6.

6.2 Data generation

Data simulated from cell death activation, synthetic biology and HIV models were performed in MATLAB R2014b using ode15s.

7 Acknowledgements

We thank Kazuhiro Aoki for discussions and providing model and data files for the MAP Kinase example. EG, KLH, and HAH acknowledge funding from the American Institute of Mathematics (AIM). EG was supported by the US National Science Foundation grant DMS-1304167. BD was partially supported by NSF DMS-1115668. KLH acknowledges support from NSF DMS-1203554. DJB gratefully acknowledges partial support from NSF DMS-1115668 and the Mathematical Biosciences Institute (MBI). HAH gratefully acknowledges funding from AMS Simons Travel Grant, EPSRC Fellowship EP/K041096/1, King Abdullah University of Science and Technology (KAUST) KUK-C1-013-04 and MPH Stumpf Leverhulme Trust Grant.

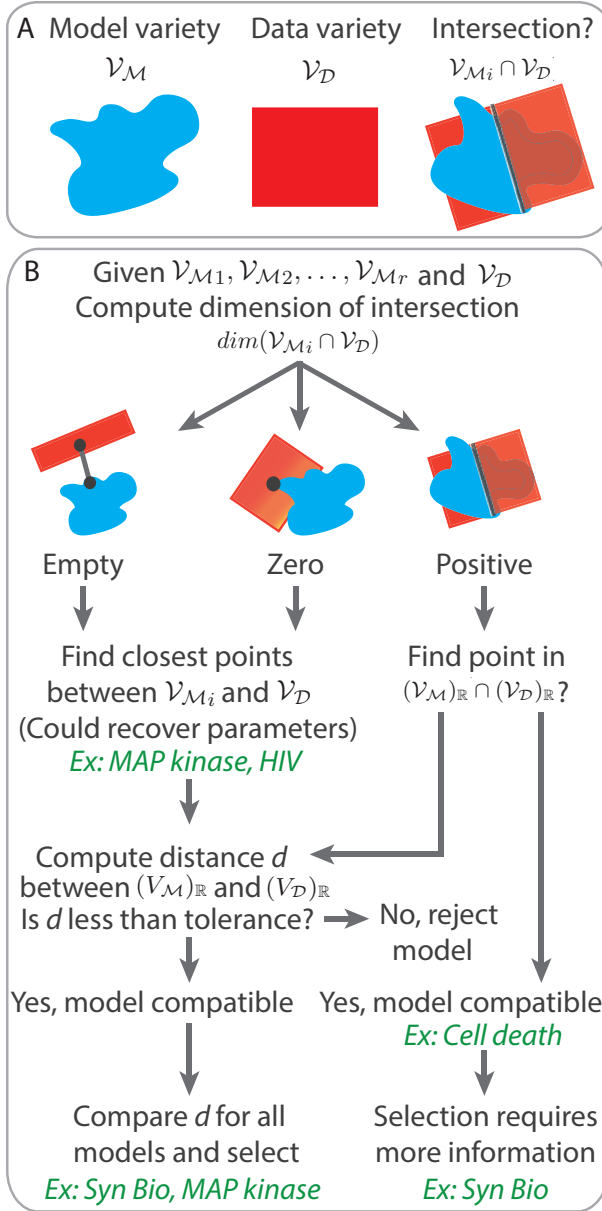


Figure S1: Schematic of numerical algebraic geometry framework corresponding to Algorithms I–III. (A) Input to algorithms include model (system of polynomials) translated into a model variety (red), and steady-state data translated into a data variety (blue). (B) Flow chart of model compatibility, parameter estimation and model selection methods. Examples (green) are described in Results section.

References

- [1] Chamberlin TC (1965) The Method of Multiple Working Hypotheses: With this method the dangers of parental affection for a favorite theory can be circumvented. *Science (New York, N.Y.)* 148:754–759.

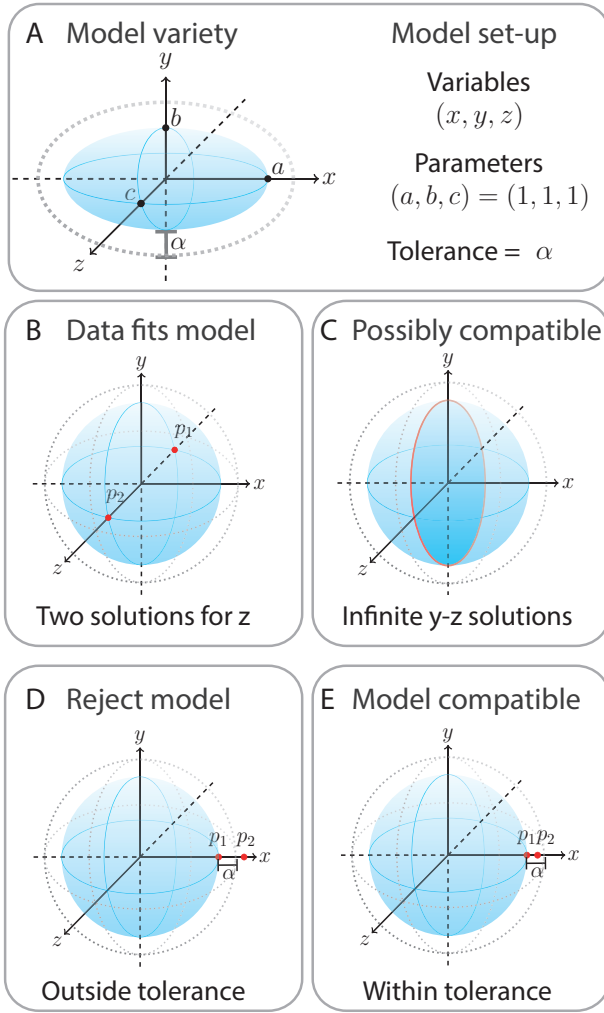


Figure S2: Simple example demonstrating model compatibility following Algorithm 1.

- [2] Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer-Verlag), 2 edition.
- [3] Kirk PDW, Thorne T, Stumpf MPH (2013) Model selection in systems and synthetic biology. *Current opinion in biotechnology* 24:767–774.
- [4] Liepe J, et al. (2014) A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature protocols* 9:439–456.
- [5] Lillacci G, Khammash M (2010) Parameter Estimation and Model Selection in Computational Biology. *PLoS Computational Biology* 6:e1000696.
- [6] Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. *Science* 220:671–680.
- [7] Černý V (1985) Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications* 45:41–51.

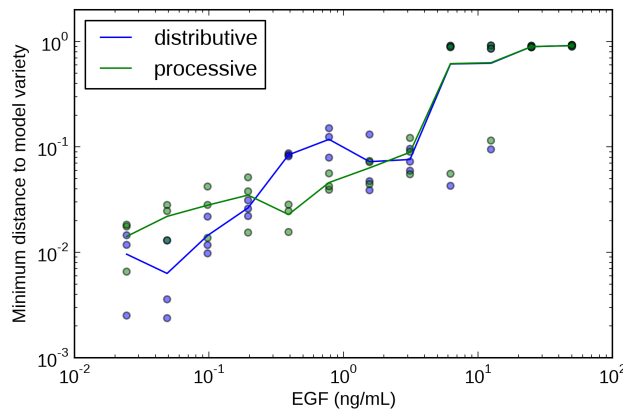


Figure S3: MAP Kinase model selection using 36 data points from 3 model variables (aggregate phosphoforms of ERK) taken in triplicate at each of 12 EGF stimulation levels.

- [8] Rumschinski P, Borchers S, Bosio S, Weismantel R, Findeisen R (2010) Set-base dynamical parameter estimation and model invalidation for biochemical reaction networks. *BMC systems biology* 4:69.
- [9] Mélykúti B, August E, Papachristodoulou A, El-Samad H (2010) Discriminating between rival biochemical network models: three approaches to optimal experiment design. *BMC systems biology* 4:38.
- [10] Anderson J, Papachristodoulou A (2009) On validation and invalidation of biological models. *BMC bioinformatics* 10:132+.
- [11] Manrai AK, Gunawardena J (2008) The Geometry of Multisite Phosphorylation. *Biophysical Journal* 95:5533–5543.
- [12] Ho KL, Harrington HA (2010) Bistability in apoptosis by receptor clustering. *PLoS Computational Biology* 6:e1000956.
- [13] Harrington HA, Ho KL, Thorne T, Stumpf MPH (2012) Parameter-free model discrimination criterion based on steady-state coplanarity. *Proceedings of the National Academy of Sciences of the United States of America* 109:15746–15751.
- [14] MacLean AL, Rosen Z, Byrne H, Harrington HA (2015) Parameter-free methods distinguish Wnt pathway models and guide design of experiments. *Proceedings of the National Academy of Sciences of the United States of America* 112:2652–2657.
- [15] Drton M, Sturmfels B, Sullivant S (2009) *Lectures on Algebraic Statistics*, Oberwolfach Seminars (Springer, Basel) Vol. 39.
- [16] Sommese AJ, Wampler CW (2005) *The numerical solution to systems of polynomials arising in engineering and science* (World Scientific, Singapore).
- [17] Bates DJ, Hauenstein JD, Sommese AJ, Wampler CW (2014) Bertini: Software for numerical algebraic geometry. available at bertini.nd.edu.
- [18] Bates DJ, et al. (2014) Comparison of probabilistic algorithms for analyzing the components of an affine algebraic variety. *Applied Mathematics and Computation* 231:619–633.
- [19] Akaike H (1971) *Information theory and an extension of the maximum likelihood principle* eds Petrov B, Csaki F (Akademiai Kiadó), pp 267–281.
- [20] Schwarz GE (1978) Estimating the dimension of a model. *Annals of Statistics* 6:461–464.

- [21] Gross E, Petrović S, Verschelde J (2013) Interfacing with phcpack. *Journal of Software for Algebra and Geometry* 5:20–25.
- [22] Hauenstein JD, Rodriguez JI, Sturmfels B (2014) Maximum likelihood for matrices with rank constraints. *Journal of Algebraic Statistics* 5:18–38.
- [23] Hauenstein JD (2013) Numerically computing real points on algebraic sets. *Acta Applicandae Mathematicae* 125:105–119.
- [24] Rouillier F, Roy MF, Din MSE (2000) Finding at least one point in each connected component of a real algebraic set defined by a single equation. *Journal of Complexity* 16:716 – 750.
- [25] Seidenberg A (1954) A new decision method for elementary algebra. *Ann. of Math.* 2:365–374.
- [26] Draisma J, Horobeț E, Ottaviani G, Sturmfels B, Thomas RR (2015) The euclidean distance degree of an algebraic variety. *Foundations of Computational Mathematics*.
- [27] Verschelde J (1999) Algorithm 795: Phcpack: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Transactions on Mathematical Software (TOMS)* 25:251–276.
- [28] Lee TL, Li TY, Tsai CH (2008) Hom4ps-2.0: a software package for solving polynomial systems by the polyhedral homotopy continuation method. *Computing* 83:109–133.
- [29] Bates DJ, Hauenstein JD, Sommese AJ, Wampler CW (2013) *Numerically solving polynomial systems with Bertini* (SIAM) Vol. 25.
- [30] Lu Y, Bates DJ, Sommese AJ, Wampler CW (2007) Finding all real points of a complex curve. *Contemporary Mathematics* 448:183.
- [31] Besana GM, Di Rocco S, Hauenstein JD, Sommese AJ, Wampler CW (2013) Cell decomposition of almost smooth real algebraic surfaces. *Numerical Algorithms* 63:645–678.
- [32] Brake D, et al. (2014) in *Mathematical Software ICMS 2014*, Lecture Notes in Computer Science, eds Hong H, Yap C (Springer Berlin Heidelberg) Vol. 8592, pp 175–182.
- [33] Thomson CB (1995) Apoptosis in the pathogenesis and treatment of disease. *Science* 267:1456–1462.
- [34] Raff M (1998) Cell suicide for beginners. *Nature* 396:119–122.
- [35] Meier P, Finch A, Evan G (2000) Apoptosis in development. *Nature* 407:796–801 0028-0836 (Print) Journal Article Review.
- [36] Fulda S, Debatin KM (2006) Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy. *Oncogene* 25:4798–4811.
- [37] Scott FL, et al. (2009 Feb 19) The fas-fadd death domain complex structure unravels signalling by receptor clustering. *Nature* 457:1019–1022.
- [38] Siegal-Gaskins D, Franco E, Zhou T, Murray RM (2014) An analytical approach to bistable biological circuit discrimination using real algebraic geometry. *BioRxiv*.
- [39] Gardner T, Cantor C, JJ C (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature* 403:339–342.
- [40] Hasty J, Isaacs F, Dolnik M, McMillen D, Collina J (2001) Designer gene networks: Towards fundamental cellular control. *Chaos* 11:207–220.
- [41] Siegal-Gaskins D, Mejia-Guerra M, Smith G, E G (2011) Emergence of switch-like behavior in a large family of simple biochemical networks. *PLoS Computational Biology* 7:e1002039.
- [42] Hernandez-Vargas EA, Middleton R, Mehta D (2011) Towards modeling hiv long term behavior. *Proceedings of the 18th IFAC World Congress* 18:581–586.

- [43] Hadjiandreu M, Conejeros R, Vassiliadis V (2007) Towards a long-term model construction for the dynamic simulation of hiv infection. *Mathematical Bioscience and Engineering* 4:489–504.
- [44] Futran AS, James A, Seger R, Shvartsman SY (2013) ERK as a Model for Systems Biology of Enzyme Kinetics in Cells. *Current Biology* 23:R972–R979.
- [45] Aoki K, Yamada M, Kunida K, Yasuda S, Matsuda M (2011) Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* 108:12675–12680.
- [46] Silk D, Kirk P, Barnes C, Toni T, Stumpf M (2014) Model selection in systems biology depends on experimental design. *PLOS Computational Biology* 10.
- [47] Gross E, Harrington HA, Rosen Z, Sturfels B (2015) Algebraic Systems Biology: A Case Study for the Wnt Pathway. *arXiv:1502.03188*.
- [48] Meshkat N, Sullivant S (2014) Identifiable reparametrizations of linear compartment models. *Journal of Symbolic Computation* 63:46–67.
- [49] Basu S, Pollack R, Roy MF (2006) *Algorithms in Real Algebraic Geometry (Algorithms and Computation in Mathematics)* (Springer-Verlag New York, Inc., Secaucus, NJ, USA).

Supporting Information for: Numerical algebraic geometry for model selection and its application to the life sciences

Elizabeth Gross¹, Brent R. Davis², Kenneth L. Ho³, Daniel J. Bates², Heather A. Harrington⁴

¹Department of Mathematics, San José State University,

²Department of Mathematics, Colorado State University,

³Department of Mathematics, Stanford University,

⁴Mathematical Institute, University of Oxford.

Contents

1	Geometry	1
1.1	Numerical algebraic geometry for isolated solutions	1
1.2	Numerical algebraic geometry for positive-dimensional solution sets	2
1.3	Software for numerical algebraic geometry	3
1.4	Geometry specific to presented algorithms	3
2	Choice of test statistics and parameter estimates	3
2.1	Maximum likelihood	3
3	Algorithm modifications	5
3.1	Solving the constrained optimization problem	5
3.2	Removing extinction components	6
4	Results	7
4.1	Cell death activation	7
4.2	Synthetic biology and experimental design	11
4.3	Epidemiology HIV	13
4.4	Multisite phosphorylation	15
4.4.1	Biology of MAP Kinase system	15
4.4.2	Mathematical models	15
4.4.3	Model selection and parameter estimation computations	18

1 Geometry

In this section we briefly review the some of the fundamental geometric concepts needed for the methods in the main text.

1.1 Numerical algebraic geometry for isolated solutions

Numerical algebraic geometry refers to the use of numerical methods, particularly homotopy continuation-based methods, to compute approximations to solutions of polynomial systems. In other words, given a polynomial system $F : \mathbb{C}^N \rightarrow \mathbb{C}^n$ with n equations in N variables, numerical algebraic geometry seeks to

find numerical approximations of all $z \in \mathbb{C}^N$ such that $F(z) = 0$. It may be the case that the solution set of F has infinitely many such points (curves, surfaces, etc.), in which case the data structure that encodes the solutions is called a *witness set*. See the description of the *numerical irreducible decomposition* below for more on this. For simplicity, we assume in this section that $N = n$. The books [1, 2] and the references therein provide much more detailed explanations than those included in this section, from more mathematical and computational perspectives.

The core technique for most numerical algebraic geometry algorithms is *homotopy continuation*. The idea of homotopy continuation is to cast the polynomial system to be solved, say $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$, as a member of a parameterized family of polynomial systems, $H : \mathbb{C}^N \times \mathbb{C} \rightarrow \mathbb{C}^N$, called a homotopy $H(z, t)$ with parameter $t \in \mathbb{C}$, that includes one polynomial system $G : \mathbb{C}^N \rightarrow \mathbb{C}^N$ that is easily solved and has the special property of having “enough” isolated solutions. In this document, we use the **Bertini** standard assumption that $H(z, 1) = G(z)$ and $H(z, 0) = F(z)$, i.e., t marches from 1 to 0. There are several canonical options for the construction of such a homotopy, and the reader is encouraged to consult the general references above and the references therein for further details.

As t varies from 1 to 0, some results from algebraic geometry tell us that the solutions of the polynomial system $H(z, t) = 0$ vary continuously and generally stay distinct until $t = 0$, where they may converge to solutions of $F(z) = 0$ or diverge. More specifically, there is a measure zero subset of $t \in \mathbb{C}$, meaning a finite set of points in this particular parameter space, over which two or more solutions coalesce. Such occurrences are thus probability zero events and, furthermore, can be detected on the fly and avoided. Said more technically, there is a *Zariski open, dense* set of the parameter space above which the solution set is finite and consists of a fixed number of solutions. Here, “Zariski” refers to the *Zariski topology*, for which basic open sets are the complements of solution sets of polynomial systems.

In practice, t is moved in discrete increments, not continuously. For each solution at $t = 1$, a path of solutions is *tracked* using numerical predictor/corrector methods as t advances to 0. Implementations typically utilize adaptive step lengths and adaptive precision. There are far too many details about this procedure to give a thorough explanation here. Instead, refer to the references above (and those therein) for further details.

Ultimately, the output of this procedure is a superset of numerical approximations of the isolated solutions of $F(z) = 0$, possibly including approximations to points lying on positive-dimensional components, if any. It is important to note that this procedure necessarily works over \mathbb{C} and finds all complex solutions. Real solutions could be buried somewhere within the complex solutions, and it is particularly difficult to extract these outside of the zero-dimensional case. However, methods do exist to extract such a real point [3–5]; here we use the method of [5], which is guaranteed (with probability one) to minimize the distance of a prescribed real point to each real connected component of the variety defined by $F(z) = 0$.

1.2 Numerical algebraic geometry for positive-dimensional solution sets

For solution sets of positive dimension (curves, surfaces, etc.), there is an extension of homotopy continuation referred to as the *numerical irreducible decomposition (NID)*. As opposed to the case of systems of linear equations (at most one solution component of one dimension), there may be many components of many different dimensions. For example, one solution set might consist of seven components of dimension four, five surfaces, three curves and 15 isolated points. Furthermore, components may be singular, meaning that the Jacobian matrix is rank-deficient throughout the component.

Technical definitions of dimension, *irreducible component*, and the like go a bit beyond the scope of this paper. It is enough to know that each “piece” of a solution set has a fixed dimension (e.g., a curve has dimension one, a surface two, etc.) and by the *dimension* of a solution set of a polynomial system of equations, we mean the maximum of the dimensions of the irreducible components.

The *numerical irreducible decomposition* of a solution set of a polynomial system consists of a catalog of the dimensions and degrees of each irreducible component, along with a set of *witness points* on each component. By *degree*, we mean the number of points in the intersection of a component with a randomly-chosen affine linear space of complementary dimension. The *witness points* on a component are then exactly these points (and thus depend on the choice of linear space). One fundamental result from algebraic geometry is that an irreducible component will *almost* always intersect a complementary-dimensional linear space exactly in a set of points *and* that the number of points is the same for *almost* any choice of linear

space. Again, this can be stated as a probability one guarantee or with Zariski open, dense sets, but we choose not to be that technical here.

1.3 Software for numerical algebraic geometry

Various numerical algebraic geometry software packages have been produced over the years. Currently, there are three main options: PHCpack [6], HOM4PS-3 [7], and Bertini [8], each with their own benefits and drawbacks. In this article, we used Bertini exclusively.

1.4 Geometry specific to presented algorithms

Theorem 1 of the main text should be familiar to those trained in multivariate calculus; this is essentially the method of Lagrange multipliers. Geometrically, this system of equations forces the normal directions of the objective function and the constraints to line up in the same direction (up to some scalar(s), the Lagrange multiplier(s)).

When working with an irreducible component of the solution set of a system of polynomial equations, it is often useful to deal with a *complete intersection*. Said simply, the idea is that computations can be more difficult if there are more equations than necessary. To see why this might be true, let us consider an example from linear algebra. Suppose we have a single linear equation in three variables, defining a plane. Now suppose we consider a system of two equations consisting of that equation and twice that equation (having the same solution set). Then the matrix of coefficients of this linear system is not full rank (not a desirable situation) and we have two equations defining a geometric object that could be described by a single equation.

In the nonlinear setting, the situation is quite similar. Having “too many” equations leads to an undesirable rank-deficient Jacobian matrix. Suppose polynomial system $F : \mathbb{C}^N \rightarrow \mathbb{C}^n$ has an irreducible component X of dimension $N - m$ (*codimension* m). Then, again with probability one, the polynomial system $A \cdot F$ has X as an irreducible component but has the “correct” number of equations, where A is a random constant matrix with n columns and m rows. Here, “correct” means the number of equations matches to codimension m . We will refer to this method as *squaring up*.

Finally for this section specific to the algorithms developed in the main text, we require the user to check two geometric facts, the meaning of which may not be entirely clear.

1. $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ refers to the intersection of the model and data varieties, as defined in the main text. To find the intersection of two solution sets, it is sufficient to simply solve the system consisting of all equations appearing in the systems for $\mathcal{V}_{\mathcal{M}}$ and $\mathcal{V}_{\mathcal{D}}$, i.e., the union of those two polynomial systems. There are more sophisticated methods, but this is sufficient.
2. The user must determine whether the intersection of $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ and $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ is empty. By this, we simply mean that one should search for real points in the intersection just described, e.g., using the method of [5].

2 Choice of test statistics and parameter estimates

In this section, we justify our procedures for model validation and parameter estimation.

2.1 Maximum likelihood

Here we justify the assertion that the test statistic given in the main text is related to likelihood maximization.

Consider first, for simplicity, the case of a single data point $\hat{y} = (\hat{y}_1, \dots, \hat{y}_m)$, which we assume is a perturbation $\hat{y} = \xi + \epsilon$ of some unknown true value $\xi = (\xi_1, \dots, \xi_m)$, where each component ϵ_i of the error $\epsilon = (\epsilon_1, \dots, \epsilon_m)$ is an independent zero-mean Gaussian random variable with variance σ_i^2 . We are interested in computing the probability that \hat{y} comes from a given model as defined by a model variety $\mathcal{V}_{\mathcal{M}}$. A point

on $\mathcal{V}_{\mathcal{M}}$ has the form (a, x, y) , where $a = (a_1, \dots, a_k)$ are the model parameter values, $x = (x_1, \dots, x_n)$ are the variable values, and $y = (y_1, \dots, y_n)$ are the outputs. The probability that \hat{y} comes from a given point $(a, x, y) \in \mathcal{V}_{\mathcal{M}}$, i.e., that \hat{y} is a perturbation of y where $(a, x, y) \in \mathcal{V}_{\mathcal{M}}$ for some a and x , is then

$$\Pr(\hat{y} \mid a, x, y) = \Pr(\hat{y} \mid \xi = y) = \prod_{i=1}^m \Pr(\hat{y}_i \mid \xi_i = y_i).$$

This is also called the likelihood $L(a, x, y)$ of (a, x, y) and we wish to find its maximizer over all $(a, x, y) \in \mathcal{V}_{\mathcal{M}}$. This can equivalently be done by considering the log-likelihood, which gives

$$\log L(a, x, y) = \sum_{i=1}^m \log \Pr(\hat{y}_i \mid \xi_i = y_i) = \sum_{i=1}^m \left(\frac{1}{2} \log 2\pi\sigma_i^2 - \frac{(\hat{y}_i - y_i)^2}{2\sigma_i^2} \right)$$

by normality. The maximizer (a^*, x^*, y^*) can therefore be found by solving the optimization problem

$$d^2 = \min_{(a, x, y) \in \mathcal{V}_{\mathcal{M}}} \sum_{i=1}^m \frac{(\hat{y}_i - y_i)^2}{\sigma_i^2},$$

where the optimum is precisely the test statistic. The values a^* , x^* , and y^* are the maximum likelihood estimates for, respectively, the parameters (estimation), the unobservable variable values (inference/recovery), and the true output values (filtering/denoising),

The test statistic d^2 itself also has a useful interpretation as follows. Suppose that \hat{y} comes from a point $(a, x, y) \in \mathcal{V}_{\mathcal{M}}$. Then

$$d^2 = \sum_{i=1}^m \frac{(\hat{y}_i - y_i^*)^2}{\sigma_i^2} \leq \sum_{i=1}^m \frac{(\hat{y}_i - y_i)^2}{\sigma_i^2}$$

by definition. But regarding each \hat{y}_i as a random variable, each term $(\hat{y}_i - y_i)/\sigma_i$ in the summation above is standard normal. Hence the right-hand side has a chi-squared distribution with m degrees of freedom (χ_m^2). The inequality should be interpreted by regarding d^2 as a random variable subject to the same source of randomness. This can be written somewhat clearer as

$$d^2(\omega) \leq \sum_{i=1}^m \frac{(\hat{y}_i(\omega) - y_i)^2}{\sigma_i^2},$$

where we have made explicit the underlying dependence of both sides on the same random realization ω . The inequality then holds for each value of ω . Consequently, we conclude that

$$\Pr(d^2 \leq u) \geq \Pr(U \leq u), \quad U \sim \chi_m^2,$$

so

$$\Pr(d^2 \geq p_\alpha) \leq \Pr(U \geq p_\alpha) = \alpha, \quad U \sim \chi_m^2,$$

where p_α is the upper α -percentile for χ_m^2 . This can be used to test the hypothesis that \hat{y} comes from $\mathcal{V}_{\mathcal{M}}$. The test statistic is also related to the log-maximum-likelihood as

$$\log L(a^*, x^*, y^*) = \frac{1}{2} \left(m \log 2\pi + \sum_{i=1}^m \log \sigma_i^2 - d^2 \right),$$

which is a useful quantity for model selection via, e.g., the Akaike or Bayesian information criteria.

Now consider the case of multiple data points $\{\hat{y}^{(j)}\}_{j=1}^p$. As before, we assume that each $\hat{y}^{(j)} = (\hat{y}_1^{(j)}, \dots, \hat{y}_m^{(j)})$ is a perturbation $\hat{y}^{(j)} = \xi^{(j)} + \epsilon^{(j)}$, where each $\epsilon_i^{(j)}$ is an independent zero-mean Gaussian random variable with variance $\sigma_{j,i}^2$. Instead of searching for one point on $\mathcal{V}_{\mathcal{M}}$, we now have to search

for p points $(a, x^{(j)}, y^{(j)})$ for $j = 1, \dots, p$ all with the same parameter values (since they come from the same fixed model realization). The probability that $\hat{y}^{(j)}$ comes from $(a, x^{(j)}, y^{(j)})$ for $j = 1, \dots, p$ is then

$$\begin{aligned} \Pr(\hat{y}^{(1)}, \dots, \hat{y}^{(p)} \mid a, x^{(1)}, y^{(1)}, \dots, x^{(p)}, y^{(p)}) &= \prod_{j=1}^p \Pr(\hat{y}^{(j)} \mid \xi^{(j)} = y^{(j)}) = \prod_{j=1}^p \prod_{i=1}^m \Pr(\hat{x}_{j,i} \mid \xi_i^{(j)} = y_i^{(j)}) \\ &\equiv L(a, x^{(1)}, y^{(1)}, \dots, x^{(p)}, y^{(p)}) \end{aligned}$$

by independence. This is essentially the same as before except that we now loop over each coordinate of each data point. Therefore, the maximum likelihood estimates $(a^*, x^{(1)*}, y^{(1)*}, \dots, x^{(p)*}, y^{(p)*})$ can be obtained by solving

$$d^2 = \min_{(a, x^{(1)}, y^{(1)}, \dots, x^{(p)}, y^{(p)}) \in \mathcal{V}_{\mathcal{M}}} \sum_{j=1}^p \sum_{i=1}^m \frac{(\hat{y}_i^{(j)} - y_i^{(j)})^2}{\sigma_{j,i}^2}.$$

The same arguments go through and we find that $\Pr(d^2 \geq p_\alpha) \leq \alpha$ for p_α the upper α -percentile for χ_{mp}^2 . The log-maximum-likelihood is related to d^2 as

$$\log L(a^*, x^{(1)*}, y^{(1)*}, \dots, x^{(p)*}, y^{(p)*}) = \frac{1}{2} \left(pm \log 2\pi + \sum_{j=1}^p \sum_{i=1}^m \log \sigma_{j,i}^2 - y \right).$$

3 Algorithm modifications

The algorithms presented in the main text are in their simplest form. Some applications require modifications, particularly if there are constraints on the variables or parameters.

3.1 Solving the constrained optimization problem

In many common settings, there exist constraints on the variable and parameter spaces. For example, in chemical reaction networks, the rate parameters are all assumed to be positive. Thus, when positivity or other constraints are present, instead of finding the weighted squared distance between two varieties, we are finding the weighted squared distance between two *semi-algebraic sets*, i.e. sets defined by polynomial equalities and inequalities as opposed to just polynomial equalities. Indeed, if we let $S_{\mathcal{M}} \subset (\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ denote the semi-algebraic set associated to the model, e.g. $S_{\mathcal{M}} = \mathcal{V}_{\mathcal{M}} \cap \mathbb{R}_{\geq 0}^{k+n+m}$, then the appropriate statistic is:

$$d^2 = \min \sum_{i=1}^m \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2} \text{ subject to } (a, x, y) \in S_{\mathcal{M}}.$$

In the case when a bound on the statistic d^2 is sufficient, then no additional work is needed. One can solve the system from Proposition 1, keeping in mind that the weighted squared distance between the closest pairs of points returned would be a lower bound on d^2 . If the closest point to $(\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$ in $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ is also an element of $S_{\mathcal{M}}$, then the squared distance would be exactly the statistic d^2 .

When the exact value of d^2 is needed, then one should solve the Karush-Kuhn-Tucker (KKT) system of equations. Let $F_1, \dots, F_r, h_1, \dots, h_s$ be polynomials in the ring $\mathbb{R}[a_1, \dots, a_k, x_1, \dots, x_n, y_1, \dots, y_m]$. Let $S_{\mathcal{M}}$ be the semi-algebraic set of all $(a, x, y) \in \mathbb{R}^{k+n+m}$ that satisfies

$$\begin{aligned} F_i(a, x, y) &= 0 \text{ for } i = 1, \dots, r \\ h_i(a, x, y) &\leq 0 \text{ for } i = 1, \dots, s \end{aligned}$$

Let $\lambda_1, \dots, \lambda_r, \mu_1, \dots, \mu_s$ be indeterminates (these are the KKT multipliers). The KKT system is

$$f = 0 \quad (1)$$

$$\lambda_1 \nabla f_1 + \dots \lambda_r \nabla f_r + \mu_1 \nabla h_1 + \dots \mu_s \nabla h_s + \begin{pmatrix} 0 \\ y - \hat{y} \end{pmatrix} = 0 \quad (2)$$

$$\mu_1 h_1 = 0 \quad (3)$$

$$\vdots \quad (4)$$

$$\mu_s h_s = 0. \quad (5)$$

In order for (a, x, y) to be a critical point, the point (a, x, y) must be a solution to this system and satisfy the inequalities $h_i \leq 0$ for $i = 1, \dots, s$ and $\mu_i \geq 0$ for $i = 1, \dots, s$. Thus, we can find the global minimum by using numerical algebraic geometry to solve the system defined by equations (1) - (5), then filtering the solutions appropriately. This combination of numerical algebraic geometry and the KKT equations was first employed in [9].

Alternatively, in some situations, it may be more efficient to minimize the objective function over $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$ and then check the boundary of $\mathcal{S}_{\mathcal{M}}$. We describe this method for when there are non-negative constraints on all indeterminates. After solving the constrained optimization problem on $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}}$, we assume, in an orderly fashion, that one of the indeterminates, say x_i , is zero. In this case, we are required to minimize the distance of $\mathcal{V}_{\mathcal{M}} \cap \{x_i = 0\}$ to $\mathcal{V}_{\mathcal{D}}$. This involves solving a smaller and related constrained optimization problem. In total, if there are N indeterminates, there are $2^N - 1$ combinations to set to zero. This amounts to solving $2^N - 1$ Lagrange multiplier systems.

One observation is that the number of systems that need to be solved explodes when N is large. Note however that the dimension of $\mathcal{V}_{\mathcal{M}} \cap \{x_i = 0\}$ is less than or equal to the dimension of $\mathcal{V}_{\mathcal{M}}$, with the inequality being strict when $V_{\mathcal{M}} \subsetneq \{x_i = 0\}$. Geometrically, there are no longer degrees of freedom in the variable x_i so the dimension is reduced. Furthermore, we can expect the dimension to be reduced by one. As we impose additional constraints, $x_j = 0$ for $i \neq j$, the dimension may drop further.

In practice, there are diminishing returns as you begin setting x_i to zero. That is, there exists some positive integer k such that $\mathcal{V}_{\mathcal{M}} \cap \{x_{i_1} = \dots = x_{i_k} = 0\}$ is zero dimensional for some indexing set $i_1 < \dots < i_k$. In this case, measuring the distance of $\mathcal{V}_{\mathcal{M}} \cap \{x_{i_1} = \dots = x_{i_k} = 0\}$ to $\mathcal{V}_{\mathcal{D}}$ using a Lagrange multiplier method is unnecessary, and, as we continue imposing additional constraints on $\mathcal{V}_{\mathcal{M}}$, either the intersection is non-empty and every observable variable is set to zero, or the model variety becomes empty. Thus, it becomes redundant or unnecessary to check additional cases.

For example, in Section 4.4.3, the MAPK model variety $\mathcal{V}_{\mathcal{M}}$ is one-dimensional and $\mathcal{V}_{\mathcal{M}} \cap \{x_i = 0\}$ becomes zero-dimensional for each i . For the cases where $x_i = x_j = 0$ for $i \neq j$, the intersection of the corresponding linear spaces with $\mathcal{V}_{\mathcal{M}}$ becomes empty. Even though there are $2^{16} - 1 = 65,535$ boundary cases to check, there are really only 16 relevant cases. See Section 4.4.3 for more explicit details on how this calculation carried out.

3.2 Removing extinction components

Given a model, it is quite common that the model variety is not irreducible but instead is the union of several irreducible components. In applications, it may be preferred to remove from consideration components that lie entirely in a coordinate hyperplane, since, in such components, one or several of the parameters and/or variables are equal to zero. For example, in a chemical reaction network, such a component is called an *extinction component* [10] since it captures the situation where one or more of the reactants have “run out.” It is common to want to avoid extinction components when estimating parameters.

Removing components where a parameter or variable is equal to zero throughout the set from consideration can be done algebraically with *saturation*. In particular, if $I_{\mathcal{M}}$ is the defining ideal of the model $\mathcal{V}_{\mathcal{M}}$ one should compute

$$I_{Main} = I_{\mathcal{M}} : (a_1 \cdots a_k \cdot x_1 \cdots x_n \cdot y_1 \cdots y_m)^{\infty} := \{f \in \mathbb{R}[a_1, \dots, a_k, x_1, \dots, x_n, y_1, \dots, y_m] : \exists k \in \mathbb{N} \text{ s.t. } (a_1 \cdots a_k \cdot x_1 \cdots x_n \cdot y_1 \cdots y_m)^k f \in I_{\mathcal{M}}\}.$$

This procedure can be performed using the `saturate` command in `Macaulay2`.

To estimate parameters such that the best estimate corresponds to a point not on an extinction component of $\mathcal{V}_{\mathcal{M}}$, one should modify Algorithm 3, replacing $\mathcal{V}_{\mathcal{M}}$ with $\mathcal{V}(I_{Main})$.

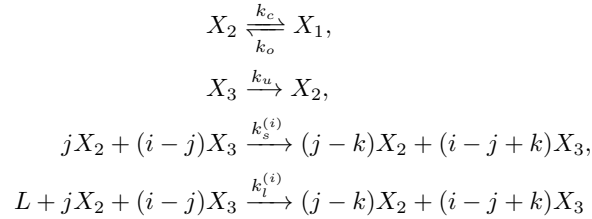
4 Results

We provide details for the calculations of examples in the main text. All code is available at: <http://www.math.sjsu.edu/~egross/NAGModelSelection/AuxillaryFiles>.

4.1 Cell death activation

We provide the details of the calculations for model compatibility of the cell death cluster model. This subsection includes detailed information regarding the solving schemes available in NAG software that were utilized. A summary for the practitioner can be found at the end of the subsection.

The model describes activation of apoptosis by death receptor Fas mechanisms [11]. The model includes constitutive receptor opening and closing, pairwise open Fas stabilization, higher-order open Fas stabilization enabled by FasL, and ligand-induced receptor opening. According to its conformational states, Fas is assumed to be one of three species: closed (X_1); open, unstable (X_2); and open, stable (X_3), i.e., active and signaling. Furthermore, let the ligand FasL be denoted by L . Then the model has the reactions



for $i \in \{2, 3\}$, $j = 1, \dots, i$, and $k = 1, \dots, j$. The first reaction describes spontaneous receptor opening and closing. The second reaction describes constitutive destabilization of open Fas. The third reaction describes cluster-stabilization by open Fas, independent of the presence of FasL. The fourth reaction describes cluster-stabilization events enabled by FasL.

Assuming mass-action kinetics, the reactions can be translated as follows:

$$\begin{cases} \dot{x}_1 &= -v_1, \\ \dot{x}_2 &= v_1 + v_2 - v_3 - v_4, \\ \dot{x}_3 &= v_3 + v_4 - v_2, \end{cases} \quad \text{where} \quad \begin{cases} v_1 &= k_o x_1 + (-k_c) x_2, \\ v_2 &= k_u x_3, \\ v_3 &= 6k_s^{(3)} x_2^3 + 3k_s^{(3)} x_2^2 x_3 + 3k_s^{(2)} x_2^2 + k_s^{(3)} x_2 x_3^2 + k_s^{(2)} x_2 x_3, \\ v_4 &= 6k_l^{(3)} x_2^3 l + 3k_l^{(3)} x_2^2 x_3 l + 3k_l^{(2)} x_2^2 l + k_l^{(3)} x_2 x_3^2 l + k_l^{(2)} x_2 x_3 l, \end{cases}$$

where v_i are the reaction velocities for the variables x_i , and lowercase letters denote the concentrations of their uppercase counterparts.

The model parameters for the cell death cluster model are

$$a = (k_o, k_c, k_u, k_s^{(2)}, k_s^{(3)}, k_l^{(2)}, k_l^{(3)})$$

and the variables are

$$x = (\ell, x_1, x_2, x_3)$$

The outputs are

$$y = (\lambda, \rho, \zeta)$$

representing, respectively, the total ligand concentration, the total receptor concentration, and the total downstream “death signal”, as given by the equations

$$\lambda - \ell = 0 \tag{6}$$

$$\rho - (x_1 + x_2 + x_3) = 0 \tag{7}$$

$$\zeta - x_3 = 0. \tag{8}$$

We set the model equations $(\dot{x}_1, \dot{x}_2, \dot{x}_3)$ to zero, and, together with equations (6)–(8), we obtain the defining equations for the model variety $\mathcal{V}_{\mathcal{M}}$. The ambient space that $\mathcal{V}_{\mathcal{M}}$ is contained in has dimension 14. This space has coordinates defined by both the model parameters a , the variables x and the outputs y .

Given an observable data point $\hat{y} = (\hat{\lambda}, \hat{\rho}, \hat{\zeta})$, we define the data variety as:

$$\mathcal{V}_{\mathcal{D}} = \{(x, a, y) \in \mathbb{C}^{14} : \lambda = \hat{\lambda}, \rho = \hat{\rho}, \zeta = \hat{\zeta}\} \quad (9)$$

for the clustering model. Note that $\mathcal{V}_{\mathcal{D}}$ has dimension 11 since there are no degrees of freedom in the variables λ, ρ , or ζ .

We first compute a numerical irreducible decomposition (NID) of $\mathcal{V}_{\mathcal{M}}$ using **Bertini**; this will aid in understanding $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$. One can verify, after computing the NID for $\mathcal{V}_{\mathcal{M}}$, that $\mathcal{V}_{\mathcal{M}}$ is a 9-dimensional complex algebraic set of degree 10 (file name: **Cluster_Model_NID**).

Now suppose we are given the following data point (taken from the model without noise):

$$\hat{y} = (\hat{\lambda}, \hat{\rho}, \hat{\zeta}) = (1.7784308, 2.31883024, 2.16896112).$$

One can then verify using the NID that $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}} \neq \emptyset$ (file name: **Cluster_Model_Data_NID**). That is, the intersection of the model variety and data variety is nonempty. Specifically, $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ is a 6-dimensional complex algebraic set of degree 5. Adding noise to the coordinates of \hat{y} taken from $\mathcal{N}(0, 0.1)$ did not affect the dimension or degree. Again, we got that $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ has dimension 6 and degree 5.

Since we are interested in model compatibility, our goal is to find at least one nonnegative point in $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$. The above computation at least provides evidence that this is the case, but it may be possible that $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ does not contain any nonnegative real points or any real points at all for that matter. We will approach this problem using the methods described in [5]. If $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ contains a real nonnegative point then we cannot reject the model and may conclude that the model is compatible with the data. If $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$ does not contain a real point then we can try and use a more general Lagrange multiplier method similar to the one employed in Section 4.4.3 in dealing with model selection.

We first randomly select a real, positive point:

$$\begin{aligned} \ell &= 6.491154749564521 \\ x_1 &= 7.317223856586703 \\ x_2 &= 6.477459631363067 \\ x_3 &= 4.509237064309449 \\ k_o &= 5.470088922863450 \\ k_c &= 2.963208056077732 \\ k_u &= 7.446928070741562 \\ k_s^{(2)} &= 1.889550150325445 \\ k_s^{(3)} &= 6.867754333653150 \\ k_l^{(2)} &= 1.835111557372697 \\ k_l^{(3)} &= 3.684845964903365 \end{aligned}$$

where each coordinate is chosen uniformly on the interval $[0, 10]$. This point will determine the observable, i.e. output, variables λ, ρ , and ζ using equations (6)–(8). Call this point $(a^*, x^*, y^*) \in \mathbb{R}^{14}$. For the time being, the left and right endpoints of each subinterval $[0, 10]$ have been chosen arbitrarily. It is unclear, at this time, how the endpoints or length of the interval affects the performance in finding nonnegative real points contained in $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$ using the methods described below.

Our aim then is to solve the constrained optimization problem:

$$\begin{aligned} \underset{y}{\text{minimize}} \quad & \|(a, x, y) - (a^*, x^*, y^*)\|^2 \\ \text{subject to} \quad & (a, x, y) \in (\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}. \end{aligned} \quad (10)$$

Geometrically, we are minimizing the distance between the chosen point (a^*, x^*, y^*) and $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$.

We will refer to the system defining $(\mathcal{V}_{\mathcal{M}}) \cap (\mathcal{V}_{\mathcal{D}})$ as $f^*(a, x, y)$. Squaring up the polynomial system will be a necessary step when utilizing the perturbed regenerative solving scheme. Squaring up was briefly described in Section 1.4 but we will give additional details here. First notice from previous computations that $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ has codimension $14 - 6 = 8$. Thus, there exists a nonempty Zariski open set $\mathcal{A} \subseteq \mathbb{C}^{8 \times 9}$ such that for every matrix $A \in \mathcal{A}$, we have $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}} \subseteq \mathcal{V}(Af^*(a, x, y))$. This means we may take 8 random \mathbb{C} -linear combinations of the equations defining $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ and, with probability one, still cut out at least $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$. It is sufficient to sample the elements of A uniformly along the complex unit circle.

As a side note, we may take additional steps to reduce complexity by replacing the matrices A with matrices of the form $[I_8 \mid b] \in \mathcal{A}$ where I_8 denotes the 8×8 identity matrix and b denotes a 8×1 column vector whose elements are sampled uniformly along the complex unit circle. This has the effect of adding a random multiple of the last function to each of the other functions. One may verify *a posteriori* if a point $x \in \mathcal{V}(Af^*(x))$ is also in $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ by function evaluation of $f^*(a, x, y)$.

The polynomial system for the optimization problem (10) is:

$$h_i(a, x, y) = 0, \text{ for } 1 \leq i \leq 8, \text{ where } h_j(a, x, y) = \sum_{k=1}^9 A_{jk} f_k^*(a, x, y) \text{ with } A_{jk} = [A]_{j,k} \quad (11)$$

$$a_i - a_i^* = \sum_{j=1}^8 \frac{\partial h_j(a, x, y)}{\partial a_i} \lambda_j, \text{ for } 1 \leq i \leq 7 \quad (12)$$

$$x_i - x_i^* = \sum_{j=1}^8 \frac{\partial h_j(a, x, y)}{\partial x_i} \lambda_j, \text{ for } 1 \leq i \leq 4 \quad (13)$$

$$y_i - y_i^* = \sum_{j=1}^8 \frac{\partial h_j(a, x, y)}{\partial y_i} \lambda_j, \text{ for } 1 \leq i \leq 3. \quad (14)$$

We call $\lambda = (\lambda_1, \dots, \lambda_8)$ the Lagrange multipliers for $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$. This is a system of 22 variables and 22 equations. Written more compactly we may write equations (11)–(14) as:

$$h(a, x, y) = 0 \quad (15)$$

$$(a, x, y)^T - (a^*, x^*, y^*)^T = J_h^T(a, x, y) \lambda^T \quad (16)$$

where $J_h(a, x, y)$ is the Jacobian matrix of $h(a, x, y)$. The Lagrange multipliers in equations (16) need not be real since we are taking \mathbb{C} -linear combinations of $f^*(a, x, y)$.

Regeneration is a numerical algebraic geometry method we found to be most relevant to solve equations (15)–(16) and is implemented in *Bertini*. Regeneration uses a linear product homotopy scheme in which each equation is built up one at a time. Depending on the ordering and structure of each equation, that can lead to huge computational savings. However, regeneration is restrictive in that it may not find all *singular* isolated solutions to equations (15)–(16). For the finer details of regeneration see reference [2].

One small adjustment we can do to solve this problem is to first solve a slightly perturbed problem (this is the strategy employed in [5]). Indeed, there is a nonempty Zariski open set Γ such that for every $\gamma \in \Gamma$ the solutions to:

$$h(a, x, y) - \gamma = 0 \quad (17)$$

$$(a, x, y)^T - (a^*, x^*, y^*)^T = J_h^T(a, x, y) \lambda^T \quad (18)$$

will be nonsingular. The benefit to first solving this system is that regeneration will now find all solutions. After the solutions to equations (17)–(18) have been computed, one can use a parameter homotopy to compute all isolated solutions (15)–(16), which may contain singular solutions.

We briefly describe the parameter homotopy employed following regeneration. The solutions of equations (17)–(18) lead to the solutions of equations (15)–(16) through a collection of homotopy paths where each homotopy path as functions of t are solutions to the straight-line homotopy function:

$$H(a, x, y, t) = \begin{cases} h(a, x, y) - t\gamma \\ (a, x, y)^T - (a^*, x^*, y^*)^T - J_h^T(a, x, y) \lambda^T \end{cases}$$

for $t \in (0, 1] \subset \mathbb{R}$. As $t \rightarrow 0$, we obtain numerical approximations to the solutions of equations (15)–(16). Additional details on parameter homotopies can be found in [1, 2]. A basic implementation of parameter

Table 1: Expected timings collected over 20 runs. The table includes the average time and standard deviations associated to the four computations described in this section.

	Timing
Compute $\mathcal{V}_{\mathcal{M}}$	0.74 sec \pm 0.09 sec
Compute $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$	0.38 sec \pm 0.04 sec
Regeneration (parallel)	6.09 sec \pm 0.61 sec
Parameter Homotopy	0.03 sec

Table 2: Nonnegative real solutions to $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$

	Solution 1	Solution 2
ℓ	1.7784308	1.7784308
x_1	0.0545838	0.0141547
x_2	0.0952853	0.1357144
x_3	2.1689611	2.1689611
λ	1.7784308	1.7784308
ρ	2.3188302	2.3188302
ζ	2.1689611	2.1689611
k_o	5.3966315	0.1924532
k_c	3.0914404	3.2734881
k_u	3.9540082	0.2856796
$k_s^{(2)}$	1.9881072	1.2768451
$k_s^{(3)}$	7.6931353	6.9985113
$k_l^{(2)}$	1.9131209	1.8363315
$k_l^{(3)}$	3.6997123	3.6848663

homotopies is found in **Bertini**. Input files for the regeneration and parameter homotopy runs may be found in the files **Cluster_Step1** and **Cluster_Step2**.

Timing summaries for the clustering model can be found in Table 1. In all cases, we have employed the use of intrinsically defined variables (see Appendix F.1.2 of [2]). These timings include computing the numerical irreducible decomposition of $\mathcal{V}_{\mathcal{M}}$, the numerical irreducible decomposition of $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$, computing the solutions to equations (17)–(18) using regeneration, and computing the solutions to (15)–(16) using the parameter homotopy. We found it most appropriate to utilize **Bertini** in serial for each computation except for regeneration which was done in parallel. Serial runs were done using a Apple MacBook Pro with 2.4 GHz Intel “Core i5” processor. Parallel runs were done using 24 (2.67 GHz Xeon-5650) compute nodes on the CentOS 5.11 operating system. In total, after reviewing the solutions, there are three solutions that correspond to real points contained in $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$. Among the three real solutions, two solutions are nonnegative. Solutions are listed in Table 2. One can verify that these are indeed solutions to $(\mathcal{V}_{\mathcal{M}})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$ by function evaluation of $f^*(a, x, y)$.

We may conclude from these computations that the clustering model $\mathcal{V}_{\mathcal{M}}$ is compatible with the observable data \hat{y} . It may be the case that there is a very large number of data points, \hat{y} , for which we would like to determine model compatibility. Fortunately, by employing a parameter homotopy scheme we may solve this problem rapidly where given each data point, the computations will be on the same order as the parameter homotopy solve in Table 1.

In summary the steps for model compatibility for the clustering model are as follows:

1. Determine the dimension of $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}$ using the NID.
2. Using the information gathered in Step 1, select a random point whose coordinates are sampled

Table 3: Each reaction described highlights whether the reaction is a forward or reversible reaction by the arrows. Here $i = 1, 2$.

monomer-dimer (\mathcal{M}_1)	dimer-dimer (\mathcal{M}_2)	single operator (\mathcal{M}_3)
$X_i \xrightarrow{k_{basi}} X_i + P_i$	$X_i \xrightarrow{k_{basi}} X_i + P_i$	$X_i \xrightarrow{k_{basi}} X_i + P_i$
$P_i \xrightarrow{k_{deg i}} \emptyset$	$P_i \xrightarrow{k_{deg i}} \emptyset$	$P_i \xrightarrow{k_{deg i}} \emptyset$
$2P_2 \xrightleftharpoons[k_{kR}]{k_{kF}} P_2P_2$	$2P_2 \xrightleftharpoons[k_{kR}]{k_{kF}} P_2P_2$	$2P_2 \xrightleftharpoons[k_{kR}]{k_{kF}} P_2P_2$
$X_1 + P_2P_2 \xrightleftharpoons[k_{nR}]{k_{nF}} X_1P_2P_2$	$X_1 + P_2P_2 \xrightleftharpoons[k_{nR}]{k_{nF}} X_1P_2P_2$	$X_2 + P_2P_2 \xrightleftharpoons[k_{qR}]{k_{qF}} X_2P_2P_2$
$X_2 + P_1 \xrightleftharpoons[k_{cR}]{k_{cF}} X_2P_1$	$X_2 + P_1P_1 \xrightleftharpoons[k_{oR}]{k_{oF}} X_2P_1P_1$	$X_2P_2P_2 \xrightarrow{k_w} X_2P_2P_2 + P_2$
	$2P_1 \xrightleftharpoons[k_{lR}]{k_{lF}} P_1P_1$	

uniformly among a nonnegative closed interval and set up equations (17)–(18).

3. Solve the perturbed equations (17)–(18) from Step 2 using a regeneration scheme, for example by setting `USEREGENERATION` to 1 in `Bertini`.
4. Solve equations (15)–(16) using a parameter homotopy. Starting solutions are among the solutions gathered in Step 3.
5. Filter the solutions gathered in Step 4 to determine if there are nonnegative real solutions.
6. We conclude that the model variety $\mathcal{V}_{\mathcal{M}}$ is compatible with the data since there is at least one solution found in Step 5.

4.2 Synthetic biology and experimental design

We demonstrate an example from synthetic biology with excess intersection ($\dim(\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}}) > 0$). We compare three bistable bio-circuits models analyzed in [12]: monomer-dimer toggle circuit (\mathcal{M}_1), dimer-dimer toggle circuit (\mathcal{M}_2), and single operator gene circuit (\mathcal{M}_3), which were initially presented in [13–15]. The model variables include concentrations of genes (X_i) and proteins (P_i) where $i = 1, 2$ as well as species complexes of the form $X_jP_iP_i, P_iP_i$.

We follow the same notation for variables and parameters as presented by [12]. The reactions governing each of the models are given in Table 3.

These variables interact following mass-action kinetics and form systems of polynomial differential equations where $\mathcal{M}_1, \mathcal{M}_2$, and \mathcal{M}_3 have 7, 8 and 6 model variables, respectively, and 10, 12, and 9 kinetic parameters, respectively. The models can be reduced by assuming that the total amount of gene 1 ($X_{1_{tot}}$) and gene 2 ($X_{2_{tot}}$) is conserved and these polynomial systems for each model are as follows. For simplicity, we use P_{11} and P_{22} for P_1P_1 and P_2P_2 , respectively.

The monomer-dimer toggle circuit (\mathcal{M}_1) system is:

$$\begin{aligned}
-k_{deg1}P_1 - k_{cF}X_2P_1 + k_{bas1}X_1 + k_{cR}(X_{2_{tot}} - X_2) &= 0 \\
-2k_{kF}P_2^2 - k_{deg2}P_2 + 2k_{kR}P_{22} + k_{bas2}X_2 &= 0 \\
k_{kF}P_2^2 - k_{kR}P_{22} - k_{nF}P_{22}X_1 + k_{nR}(X_{1_{tot}} - X_1) &= 0 \\
k_{nR}(X_{1_{tot}} - X_1) - k_{nF}P_{22}X_1 &= 0 \\
k_{cR}(X_{2_{tot}} - X_2) - k_{cF}P_1X_2 &= 0.
\end{aligned}$$

The model dimer-dimer toggle circuit (\mathcal{M}_2) system is:

$$\begin{aligned}
-2k_{iF}P_1^2 - k_{deg1}P_1 + 2k_{iR}P_{11} + k_{bas1}X_1 &= 0 \\
k_{iF}P_1^2 - k_{iR}P_{11} - k_{oF}P_{11}X_2 + k_{oR}(X_{2_{tot}} - X_2) &= 0 \\
-2k_{kF}P_2^2 - k_{deg2}P_2 + 2k_{kR}P_{22} + k_{bas2}X_2 &= 0 \\
k_{kF}P_2^2 - k_{kR}P_{22} + k_{nF}P_{22}X_1 + k_{nR}(X_{1_{tot}} - X_1) &= 0 \\
k_{nR}(X_{1_{tot}} - X_1) - k_{nF}P_{22}X_1 &= 0 \\
k_{oR}(X_{2_{tot}} - X_2) - k_{oF}P_{11}X_2 &= 0.
\end{aligned}$$

The model single-operator positive feedback circuit (\mathcal{M}_3) system is:

$$\begin{aligned}
k_{bas2}X_2 - k_{deg2}P_2 - 2k_{kF}P_2^2 + 2k_{kR}P_{22} + k_w(X_{2_{tot}} - X_2) &= 0 \\
k_{kF}P_2^2 - k_{kR}P_{22} - k_{qF}P_{22}X_2 + k_{qR}(X_{2_{tot}} - X_2) &= 0 \\
k_{qR}(X_{2_{tot}} - X_2) - k_{qF}P_{22}X_2 &= 0
\end{aligned}$$

In this example, we suppose that the total amounts $X_{1_{tot}}$ and $X_{2_{tot}}$ and specific protein synthesis and degradation parameters k_{bas1} , k_{bas2} , k_{deg1} , and k_{deg2} are known and we assume that our data are measurements of P_1 , P_2 , and their complexes P_{11} , and P_{22} , i.e. $y = (P_1, P_2, P_{11}, P_{22})$. The aim is to select the best model \mathcal{M}_1 , \mathcal{M}_2 , and \mathcal{M}_3 given the data. We simulate steady-state data $(P_1, P_2, P_{11}, P_{22}) = (0.4224, 2.4153, 0.9022, 0.4758)$ from the dimer-dimer toggle model (\mathcal{M}_2) using the following parameter and variable values:

parameter	value	parameter	value
$X_{1_{tot}}$	1.2099	k_{nF}	1.3566
$X_{2_{tot}}$	2.0660	k_{nR}	0.6521
k_{bas1}	0.8718	k_{oF}	1.5169
k_{bas2}	1.6930	k_{oR}	1.0661
k_{deg1}	1.2550	k_{iF}	3.3169
k_{deg2}	0.6341	k_{iR}	0.6559
k_{kF}	0.6580	k_{qF}	0.5057
k_{kR}	8.0681	k_{qR}	0.4844
k_{cF}	0.4675	k_w	0.1478
k_{cR}	1.1636		

We add Gaussian noise from $\mathcal{N}(0, 0.1)$ and then find $\dim(\mathcal{V}_{\mathcal{M}_i} \cap \mathcal{V}_{\mathcal{D}})$ for $i = 1, 2, 3$. We can compute the dimension of each intersection using the `dim` command in Macaulay2 or by computing a numerical irreducible decomposition in `Bertini`; we find:

$$\begin{aligned}
\dim(\mathcal{V}_{\mathcal{M}_1} \cap (\mathcal{V}_{\mathcal{D}})) &= 3, \\
\dim(\mathcal{V}_{\mathcal{M}_2} \cap \mathcal{V}_{\mathcal{D}}) &= 4, \\
\dim(\mathcal{V}_{\mathcal{M}_3} \cap (\mathcal{V}_{\mathcal{D}})) &= 3.
\end{aligned}$$

Some further computations are required to find $\dim((\mathcal{V}_{\mathcal{M}_i})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}})$. Specifically, we need to find real points in each intersection and determine whether or not those points are smooth. Computing the dimension of the real part of the intersections is more work than necessary for Algorithm 2, however, it provides an illustrative example on how to work with real varieties and the algorithm in [5].

Let $f^{(i)} = 0$ be the polynomial system defining $\mathcal{V}_{\mathcal{M}_i} \cap \mathcal{V}_{\mathcal{D}}$ for $i = 1, 2, 3$ and let $w^{(1)} \in \mathbb{R}^{17}$, $w^{(2)} \in \mathbb{R}^{20}$, $w^{(3)} \in \mathbb{R}^{15}$ be random points. Let $x^{(i)}$ be the vector of indeterminates (unknown parameters and variables) for i th model, and let c_i be the codimension of $\mathcal{V}_{\mathcal{M}_i} \cap \mathcal{V}_{\mathcal{D}}$. We can find a real point on every component of each intersection, by solving the system:

$$f^{(i)} = 0, \tag{19}$$

$$\lambda_1 \nabla f_1^{(i)} + \dots \lambda_{c_i} \nabla f_{c_i}^{(i)} + (x^{(i)} - w^{(i)}) = 0. \tag{20}$$

This is a simplified version of the system in Theorem 5 from [5]. As a remark, notice the similarity of the (19)-(20) to the system in Theorem 1. Algorithms for finding real points on a variety have been built on algorithms for minimizing the distance between a point and a variety since [16].

Once we have a real point on every component of $\mathcal{V}_{\mathcal{M}_i} \cap \mathcal{V}_{\mathcal{D}}$, we can quickly determine $\dim(\mathcal{V}_{\mathcal{M}_i})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}$ if those real points are smooth. Indeed, if \mathcal{V} is an irreducible variety, then $\dim \mathcal{V} = \dim \mathcal{V}_{\mathbb{R}}$ if \mathcal{V} contains a real smooth point (see [2, §14.1]). Checking whether a real point is smooth can be done by evaluating the Jacobian $\mathcal{V}_{\mathcal{M}_i} \cap \mathcal{V}_{\mathcal{D}}$ at the point; if the Jacobian has full rank, then the point is smooth. In our case, for the three models, every point we find is smooth and thus we are able to reach the conclusion:

$$\begin{aligned}\dim((\mathcal{V}_{\mathcal{M}_1})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}) &= 3, \\ \dim((\mathcal{V}_{\mathcal{M}_2})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}) &= 4, \\ \dim((\mathcal{V}_{\mathcal{M}_3})_{\mathbb{R}} \cap (\mathcal{V}_{\mathcal{D}})_{\mathbb{R}}) &= 3.\end{aligned}$$

The dimension analysis of the varieties $\mathcal{V}_{\mathcal{M}_i} \cap \mathcal{V}_{\mathcal{D}}$ informs us about the minimum number of additional variable and parameter values that must be measured to ensure $\mathcal{V}_{\mathcal{M}} \cap \mathcal{V}_{\mathcal{D}} = \emptyset$. For \mathcal{M}_1 we need to know at least 4 more variable and/or parameter values, for \mathcal{M}_2 we need to know at least 5 more, and for \mathcal{M}_3 we need to know at least 4 more. Thus for the remainder of the example, we assume that we the parameters k_{cF} , k_{cR} , k_{nF} , and k_{kF} are known in \mathcal{M}_1 , the parameters k_{kF} , k_{nF} , k_{iF} , k_{oF} , and k_{oR} are known in \mathcal{M}_2 , and the parameters k_{kF} , k_{qF} , k_{qR} and k_w are known in \mathcal{M}_3 . The model \mathcal{M}_3 is an example where the number of additional parameters and/or variables that need to be known/measured exceeds the amount predicted by the dimension analysis.

Now that all the intersections are empty, we run Algorithm 2, using the regeneration methods in Bertini to solve the systems resulting from Theorem 1. We find that the sum of squares (Eq. (3.1)) for each model are as follows: $d_1^2 = 2.116$, $d_2^2 = 0.000124$, and $d_3^2 = 0.6333$. Therefore, we select the \mathcal{M}_2 model, which matches the model from which the data was generated.

Solving the zero-dimensional system for the monomer-dimer toggle circuit, \mathcal{M}_1 , took 1 minute and 3 seconds on an Apple MacBook Pro with a 2.6 GHz Intel Core i5 processor. Solving the system for the dimer-dimer toggle circuit, \mathcal{M}_2 , took 1 minute and 43 seconds, and solving the system for the single-operator positive feedback circuit, \mathcal{M}_3 took 0.092 seconds.

4.3 Epidemiology HIV

To demonstrate parameter estimation we use a model that includes long-term HIV dynamics from initial virus, latency, and virus increase [17], based on [18]. Model variables x are uninfected CD4+T cells (T), infected CD4+ T cells (T_i), uninfected macrophages (M), infected macrophages (M_i), and HIV virus population (V). The reactions are considered for this model are shown in Table 4.

From these reactions, the dynamics are described by the following equations:

$$\begin{aligned}\dot{T} &= s_1 + k_1TV - k_2TV - \delta_1T \\ \dot{T}_i &= k_2TV - \delta_2T_i \\ \dot{M} &= s_2 + k_3MV - k_4MV - \delta_3M \\ \dot{M}_i &= k_4MV - \delta_4M_i \\ \dot{V} &= k_5T_i + k_6M_i - \delta_5V\end{aligned}$$

Using Macaulay2, we find that the model variety $\mathcal{V}_{\mathcal{M}}$ has two irreducible components, the main component \mathcal{V}_1 defined by the ideal

$$\begin{aligned}I_1 = \langle &5742M - 2453M_i - 130500, 259908T_i - 46607M_i + 4840000\delta_5 - 20200500, \\ &17721T + 46607M_i - 4840000\delta_5 + 2479500, 484000V\delta_5 - 184547M_i + 4840000\delta_5 - 20200500, \\ &2453M_iV - 72600M_i + 130500V \rangle\end{aligned}$$

Table 4: Reactions for HIV model. The published parameter value is used from [17], see references therein.

Description	Reaction	Parameter value
Generation of new CD4+T cells	$\emptyset \xrightarrow{s_1} T$	10
Generation of new macrophages	$\emptyset \xrightarrow{s_2} M$	1.5×10^{-1}
Proliferation of T cells by presence of pathogen	$T + V \xrightarrow{k_1} (T + V) + T$	2×10^{-3}
Infection of T cells by HIV	$T + V \xrightarrow{k_2} T_i$	3×10^{-3}
Proliferation of M by presence of pathogen	$M + V \xrightarrow{k_3} (M + V) + M$	7.45×10^{-4}
Infection of M by HIV	$M + V \xrightarrow{k_4} M_i$	5.22×10^{-4}
Proliferation of HIV within CD4+T cell	$T_i \xrightarrow{k_5} V + T_i$	5.37×10^{-1}
Proliferation of HIV within macrophage	$M_i \xrightarrow{k_6} V + M_i$	2.85×10^{-1}
Natural death of CD4+T cells	$T \xrightarrow{\delta_1} \emptyset$	0.01
Natural death of infected T cells	$T_i \xrightarrow{\delta_2} \emptyset$	0.44
Natural death of macrophages	$M \xrightarrow{\delta_3} \emptyset$	6.6×10^{-3}
Natural death of infected macrophages	$M_i \xrightarrow{\delta_4} \emptyset$	6.6×10^{-3}
Natural death of HIV	$V \xrightarrow{\delta_5} \emptyset$	3

and an extinction component V_2 defined by the ideal

$$I_2 = \langle V, M_i, 11M - 250, T_i, T - 1000 \rangle$$

We estimated the natural death of the virus, parameter δ_5 , using Algorithm 3 with main component \mathcal{V}_1 in place of \mathcal{V}_M (see Section 3.2). For \mathcal{V}_D , we used the long-term non-progressors steady-state value (Table 3 of [17]) and added noise to each variable $\sim \mathcal{N}(0, 1)$. In particular, the data variety \mathcal{V}_D is defined by the equations:

$$\begin{aligned} T - \frac{6383}{20} &= 0 \\ T_i - \frac{937}{20} &= 0 \\ M - \frac{8109}{100} &= 0 \\ M_i - \frac{13667}{100} &= 0 \\ V - \frac{2121}{100} &= 0 \end{aligned}$$

For $s_1, s_2, k_1, \dots, k_6, \delta_1, \dots, \delta_4$, we treated these parameters as known using the values from Table 1 of [17].

The varieties \mathcal{V}_1 and \mathcal{V}_D do not intersect, which we can confirm with **Macaulay2**. Using **Bertini**, we

solve the following system from Theorem 1:

$$\begin{aligned}
5742M - 2453M_i - 130500 &= 0 \\
259908T_i - 46607M_i + 4840000\delta_5 - 20200500 &= 0 \\
17721T + 46607M_i - 4840000\delta_5 + 2479500, 484000V\delta_5 - 184547M_i + 4840000\delta_5 - 20200500 &= 0 \\
2453M_iV - 72600M_i + 130500V &= 0 \\
T + 17721\lambda_3 - 6383/20 &= 0 \\
T_i + 259908\lambda_2 - 937/20 &= 0 \\
M + 5742\lambda_1 - 8109/100 &= 0 \\
2453\lambda_5 + M_i - 2453\lambda_1 - 46607\lambda_2 + 46607\lambda_3 - 184547\lambda_4 - 72600\lambda_5 - 13667/100 &= 0 \\
484000\delta_5\lambda_4 + 2453M_i\lambda_5 + V + 130500\lambda_5 - 2121/100 &= 0 \\
484000V\lambda_4 + 4840000\lambda_2 - 4840000\lambda_3 + 4840000\lambda_4 &= 0
\end{aligned}$$

There are 16 complex solutions to this equation, 3 of which are real. The real point resulting in the smallest sum of squared errors $d^2 = 0.2884$ is:

$$\begin{aligned}
(T, T_i, M, M_i, V, \delta_5, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = \\
(319.408, 46.404, 81.1544, 136.767, 21.3079, 2.99876, \\
-0.0000112074, 0.00000171594, -0.0000145481, -0.00000519486, 0.0000159701)
\end{aligned}$$

Thus, Algorithm 3 returns $\bar{\delta}_5 = 2.99876$, which we can compare to the true value $\delta_5 = 3$.

4.4 Multisite phosphorylation

Here we describe the details for the multisite phosphorylation model selection and parameter estimation computations. First we describe the relevant biology, next we present the mathematical models of the distributive and processive mechanisms, then we apply our model selection method using data from [19]. We also estimate the relationship between the EGF concentration and activation of the pathway described by the parameter k_1 (see Table 11).

4.4.1 Biology of MAP Kinase system

Many cellular decisions are governed by molecular post-translational modifications. One type of modification, phosphorylation, is the addition of a phosphate group to a site of a substrate by an enzyme called a kinase. Some proteins (substrates) require multiple phosphate groups to be added by the kinase before the protein is activated/de-activated by these modifications. One well-studied signaling system is the MAP Kinase pathway, with kinase MEK and its substrate ERK; however the mechanism by which the phosphate group is added has been debated. Either MEK could phosphorylate ERK, disassociate and then phosphorylate again, called *distributive*; or MEK could bind and phosphorylate in sequence, called *processive*. Aoki et al [19] showed experimentally (with mathematical models) that this mechanism is different *in vitro* than *in vivo*. This experiment included 12 different levels of EGF stimulus ranging from 0.0244140625 ng/mL to 50 ng/mL. EGF activates cRAF which then phosphorylates MEK and finally doubly phosphorylates ERK. The data are measurements of three replicates of nonphosphorylated ERK (np-ERK), tyrosine monophosphorylated ERK (pY-ERK), and doubly phosphorylated ERK (pTpY-ERK), at each stimulus level. These data are given as percentage of total ERK (ERK), so we use the concentration measurement for each of these ERK states.

4.4.2 Mathematical models

The model variables and parameters are given in Table 5. The model parameters for the distributive model are

$$a = (k_1, \dots, k_{27}, c_1, c_2),$$

Table 5: Description of variables and parameters for distributive and processive MAP Kinase models

variable	species	parameter	name	parameter	name
x_1	MEK	k_1	kphos_MEK_pMEK	k_{15}	kdphos_pY_np_cyt
x_2	cRAF	k_2	kdphos_pMEK_MEK	k_{16}	kdphos_pT_np_cyt
x_3	pMEK	k_3	kf_MEK_ERK_binding	k_{17}	kdphos_pTpY_pY_nuc
x_4	np-ERK_cyt	k_4	kb_MEK_ERK_dissociation	k_{18}	kdphos_pTpY_pT_nuc
x_5	MEK_np-ERK	k_5	kimport_np	k_{19}	kdphos_pY_np_nuc
x_6	np-ERK_nuc	k_6	kexport_np	k_{20}	kdphos_pT_np_nuc
x_7	pY-ERK_cyt	k_7	kimport_pY	k_{21}	kphos_np_pY
x_8	pY-ERK_nuc	k_8	kexport_pY	k_{22}	kphos_pY_pTpY
x_9	pT-ERK_cyt	k_9	kimport_pT	k_{23}	kphos_pT_pTpY
x_{10}	pT-ERK_nuc	k_{10}	kexport_pT	k_{24}	kf_MEK_ERK_binding
x_{11}	pTpY-ERK_cyt	k_{11}	kimport_pTpY	k_{25}	kb_MEK_ERK_dissociation
x_{12}	pTpY-ERK_nuc	k_{12}	kexport_pTpY	k_{26}	kphos_np_pY
x_{13}	pMEK_np-ERK	k_{13}	kdphos_pTpY_pY_cyt	k_{27}	kphos_pY_pTpY_MEKERK
x_{14}	pMEK_pY-ERK	k_{14}	kdphos_pTpY_pT_cyt	c_2, c_1	cyt_vol, nuc_vol

Table 6: Reaction velocities for the MAP Kinase distributive and processive model. The processive model uses the additional reaction velocities v_{18}, v_{19}, v_{20} .

$v_1 = k_1 x_1 x_2 - k_2 x_3$	$v_2 = k_3 x_1 x_4 - k_4 x_5$	$v_3 = k_5 x_4 - c_2 k_6 x_6$
$v_4 = k_7 x_7 - c_2 k_8 x_8$	$v_5 = k_9 x_9 - c_2 k_{10} x_{10}$	$v_6 = k_{11} x_{11} - c_2 k_{12} x_{12}$
$v_7 = k_{13} x_{11}$	$v_8 = k_{14} x_{11}$	$v_9 = k_{15} x_7$
$v_{10} = k_{16} x_9$	$v_{11} = c_2 k_{17} x_{12}$	$v_{12} = c_2 k_{18} x_{12}$
$v_{13} = c_2 k_{19} x_8$	$v_{14} = c_2 k_{20} x_{10}$	$v_{15} = k_{21} x_3 x_4$
$v_{16} = k_{22} x_3 x_7$	$v_{17} = k_{23} x_3 x_9$	
$v_{18} = k_{24} x_3 x_4 - k_{25} x_{13}$	$v_{19} = k_{26} x_{13}$	$v_{20} = k_{27} x_{14}$

the variables are

$$x = (x_1, \dots, x_{12}, \text{cRAF}_{tot}, \text{MEK}_{tot}, \text{ERK}_{tot}),$$

and the outputs are

$$y = (\text{np-ERK}, \text{pY-ERK}, \text{pYpT-ERK}).$$

The variables for the processive model are the same as for the distributive model except for two additional variables x_{13}, x_{14} . The reaction velocities are given in Table 6 and the corresponding equations are given in Table 7. Note in Table 7 that there are various conserved species concentrations in addition to the ODEs.

We use the *in vitro* parameters estimates from Table S2 in reference [19] for $k_2, \dots, k_{27}, c_1, c_2$ and the conserved quantities $\text{MEK}_{tot}, \text{cRAF}_{tot}, \text{ERK}_{tot}$, as given in Table 8. The remaining parameter, k_1 , describes the rate of MEK phosphorylation and depends on the level of EGF stimulation, which varies throughout the data. Thus, we left it as a free variable and estimated it as a byproduct of distance minimization (11). The output variables are np-ERK, pY-ERK, and pYpT-ERK, which are sums of species concentrations. For the distributive model, the output equations are:

$$\text{np-ERK} - (x_4 + x_5 + x_6) = 0 \quad (21)$$

$$\text{pY-ERK} - (x_7 + x_8) = 0 \quad (22)$$

$$\text{pYpT-ERK} - (x_{11} + x_{12}) = 0 \quad (23)$$

Table 7: Equations for distributive and processive MAP Kinase models

Variable	Distributive	Processive
$\dot{x}_1 =$	$-v_1 - v_2$	$-v_1 - v_2$
$\dot{x}_2 =$	0	0
$\dot{x}_3 =$	v_1	$v_1 - v_{18} + v_{20}$
$\dot{x}_4 =$	$-v_2 - v_3 + v_9 + v_{10} - v_{15}$	$-v_2 - v_3 + v_9 + v_{10} - v_{18}$
$\dot{x}_5 =$	v_2	v_2
$\dot{x}_6 =$	$v_3 + v_{13} + v_{14}$	$v_3 + v_{13} + v_{14}$
$\dot{x}_7 =$	$-v_4 + v_7 - v_9 + v_{15} - v_{16}$	$-v_4 + v_7 - v_9 - v_{16}$
$\dot{x}_8 =$	$v_4 + v_{11} - v_{13}$	$v_4 + v_{11} - v_{13}$
$\dot{x}_9 =$	$-v_5 + v_8 - v_{10} - v_{17}$	$-v_5 + v_8 - v_{10} - v_{17}$
$\dot{x}_{10} =$	$v_5 + v_{12} - v_{14}$	$v_5 + v_{12} - v_{14}$
$\dot{x}_{11} =$	$-v_6 - v_7 - v_8 + v_{16} + v_{17}$	$-v_6 - v_7 - v_8 + v_{16} + v_{17} + v_{20}$
$\dot{x}_{12} =$	$v_6 - v_{11} - v_{12}$	$v_6 - v_{11} - v_{12}$
$\dot{x}_{13} =$		$v_{18} - v_{19}$
$\dot{x}_{14} =$		$v_{19} - v_{20}$
0 =	$\text{MEK}_{tot} - (x_1 + x_3 + x_5)$	$\text{MEK}_{tot} - (x_1 + x_3 + x_5 + x_{13} + x_{14})$
0 =	$\text{cRAF}_{tot} - x_2$	$\text{cRAF}_{tot} - x_2$
0 =	$\text{ERK}_{tot} - \sum_{i=4}^{12} x_i$	$\text{ERK}_{tot} - \sum_{i=4}^{14} x_i$

Table 8: Parameter values for MAP Kinase models

parameter	value	parameter	value	parameter	value
k_2	0.0096	k_{13}	0.004	k_{24}	0.18
k_3	0.18	k_{14}	0.0055	k_{25}	0.27
k_4	0.27	k_{15}	0.0067	k_{26}	0.073
k_5	0.0017	k_{16}	0.0068	k_{27}	0.05
k_6	0.013	k_{17}	0.0032	c_1	1.0
k_7	0.0025	k_{18}	0.0038	c_2	0.2
k_8	0.017	k_{19}	0.0077	cRAF_{tot}	0.013
k_9	0.0022	k_{20}	0.0058	MEK_{tot}	1.2
k_{10}	0.049	k_{21}	0.039	ERK_{tot}	0.74
k_{11}	0.0082	k_{22}	0.021		
k_{12}	0.0076	k_{23}	0.02		

whereas for the processive model, we include two additional species, so the output equations become:

$$\text{np-ERK} - (x_4 + x_5 + x_6 + x_{13}) = 0 \quad (24)$$

$$\text{pY-ERK} - (x_7 + x_8 + x_{14}) = 0 \quad (25)$$

$$\text{pYpT-ERK} - (x_{11} + x_{12}) = 0. \quad (26)$$

4.4.3 Model selection and parameter estimation computations

The model variety $\mathcal{V}_{\mathcal{M}_d}$ of the distributive model is defined by (21) - (23) and the equations obtained by setting the “Distributive” column of Table 7 equal to zero. We will refer to the system defining $\mathcal{V}_{\mathcal{M}_d}$ as $F = 0$. The model variety $\mathcal{V}_{\mathcal{M}_p}$ of the processive model is defined by (24) - (26) and the equations obtained by setting the “Processive” column of Table 7 equal to zero. The ambient dimension of $\mathcal{V}_{\mathcal{M}_d}$ is 16 since the coordinates that define $\mathcal{V}_{\mathcal{M}_d}$ include x_1, \dots, x_{12} , np-ERK, pY-ERK, pYpT-ERK, and the model parameter k_1 ; all other parameters and variables we treat as known constants. Similarly the ambient dimension for the processive model is 18 as we include the additional variables x_{13} and x_{14} .

Given data $\hat{y} = (\widehat{\text{np-ERK}}, \widehat{\text{pY-ERK}}, \widehat{\text{pYpT-ERK}})$ we define the data variety for the distributive model as:

$$\mathcal{V}_{\mathcal{D}_d} = \{(a, x, y) \in \mathbb{C}^{16} : y = \hat{y}\}$$

The data variety $\mathcal{V}_{\mathcal{D}_d}$ has dimension 13. The specific data that we used may be found in the supplementary data file (`aoki_data.txt`). The data variety $\mathcal{V}_{\mathcal{D}_p}$ for the processive model is defined similarly.

The computations that follow will be for the distributive model. The computations for the processive model will be nearly identical so we do not describe them in the same level of detail. When the results are discussed we will be sure to record information for both models.

We first compute a numerical irreducible decomposition (NID) of $\mathcal{V}_{\mathcal{M}_d}$ using the `Bertini.m2` [20] the `Macaulay2` interface for `Bertini` to solve [8, 20, 21]. With the NID, one can verify that $\mathcal{V}_{\mathcal{M}_d}$ is a one-dimensional complex algebraic set of degree 8 (filename: `MAPK_D_Model_NID`). Similarly for the processive model, the model variety $\mathcal{V}_{\mathcal{M}_p}$ is a one-dimensional complex algebraic set of degree 11 (filename: `MAPK_P_Model_NID`). There are several variables that may be intrinsically defined to save computation. For example, x_1, x_2, x_7 , and x_{11} can be written in terms of the other variables followed by x_4 . One may also verify that $\mathcal{V}_{\mathcal{M}_d} \cap \mathcal{V}_{\mathcal{D}_d} = \emptyset$ using `Bertini` (filename: `MAPK_D_Model_Data_NID`). Here, one can also define the variables np-ERK, pY-ERK, and pYpT-ERK intrinsically to save computation. Since $\mathcal{V}_{\mathcal{M}_d} \cap \mathcal{V}_{\mathcal{D}_d} = \emptyset$, Algorithm 2 instructs us to minimize the distance between $(\mathcal{V}_{\mathcal{M}_d})_{\mathbb{R}}$ and $(\mathcal{V}_{\mathcal{D}_d})_{\mathbb{R}}$.

Squaring up the polynomial system defining $\mathcal{V}_{\mathcal{M}_d}$ will be a necessary step to construct the polynomial system from Proposition 1. This procedure was described briefly in Section 1.4 and in more detail in Section 4.1. The codimension of $\mathcal{V}_{\mathcal{M}_d}$ is $c = 16 - 1 = 15$, the dimension of the ambient space minus the dimension of $\mathcal{V}_{\mathcal{M}_d}$ as determined by the numerical irreducible decomposition. Let $A \in \mathbb{C}^{15 \times 17}$ whose entries are taken randomly from the complex unit circle. The polynomial system from Proposition 1 becomes:

$$f'_j(a, x, y) = 0, \text{ for } 1 \leq i \leq 15, \text{ where } f'_j(x) = \sum_{k=1}^{17} A_{jk} F_k(a, x, y) \text{ with } A_{jk} = [A]_{j,k} \quad (27)$$

$$0 = \sum_{j=1}^{15} \frac{\partial f'_j(a, x, y)}{\partial a_1} \lambda_j, \quad (28)$$

$$0 = \sum_{j=1}^{15} \frac{\partial f'_j(a, x, y)}{\partial x_i} \lambda_j, \text{ for } 1 \leq i \leq 12 \quad (29)$$

$$y_i - \hat{y}_i = \sum_{j=1}^{15} \frac{\partial f'_j(a, x, y)}{\partial y_i} \lambda_j, \text{ for } 1 \leq i \leq 3 \quad (30)$$

The variables $\lambda = (\lambda_1, \dots, \lambda_{15})$ are the Lagrange multipliers. This is a system of 31 variables and 31 equations. We collect the solutions $(a, x, y, \lambda) \in \mathbb{R} \times \mathbb{R}^{12} \times \mathbb{R}^3 \times \mathbb{C}^{15}$ to equations (27)–(30) and then compute $\|y - \hat{y}\|_2^2$ for each solution.

In Section 3.1, we explained an issue that can arise in solving constrained optimization problems such as ones arising from problem (1)–(5). In this example, we want to ensure that $x_1, \dots, x_{12}, a_1, y_1, y_2, y_3$ are non-negative, i.e. $\mathcal{S}_{\mathcal{M}_d} = \mathcal{V}_{\mathcal{M}_d} \cap \mathbb{R}_{\geq 0}^{16}$. To minimize the distance between $\mathcal{S}_{\mathcal{M}_d}$ and $\mathcal{V}_{\mathcal{D}_d}$ using a numerical

Table 9: Path counts on processive and distributive models. ‘ $\{(a, x, y), \lambda\}$ -hom’ corresponds to a $\{(a, x, y, \lambda)\}$ -homogeneous variable grouping and ‘intrinsic x_2 ’ corresponds to the system where x_2 is intrinsically defined.

	Total Degree	$\{(a, x, y), \lambda\}$ -hom	$\{(a, x, y), \lambda\}$ -hom + intrinsic x_2
Processive Model	124416 paths	3744 paths	1152 paths
Distributive Model	248832 paths	7488 paths	2304 paths

algebraic geometry approach, we should solve the system (27)–(30), and then solve the system again 16 more times, setting one of $x_1, \dots, x_{12}, a_1, y_1, y_2, y_3$ to zero each time. We know that this will be sufficient to check the boundary, since the intersection of $\mathcal{V}_{\mathcal{M}_d}$ with any coordinate hyperplane is zero dimensional.

The data set we used can be found in the file (`aokidata.txt`). There are 36 sets of data points where each set consists of a triple $\hat{y} = (\widehat{\text{np-ERK}}, \widehat{\text{pY-ERK}}, \widehat{\text{pYpT-ERK}})$. Each data point defines a data variety, which we will denote $(V_{D_i})_{\mathbb{R}}$ for the i th data point.

The polynomial system (27)–(30) is a classic example of a parameterized system of polynomial equations with parameter \hat{y} (notice that this is different than the rate parameters of the ODE). Let $J_{f'}^T(a, x, y)$ be the Jacobian matrix of $f'(a, x, y)$. The theory for parameter homotopies states that there is a nonempty Zariski open set $U \subset \mathbb{C}^3$ such that for every $\hat{y}^* \in U$ the nonsingular solutions of:

$$f'(a, x, y) = 0 \quad (31)$$

$$\begin{bmatrix} 0 \\ y^T - (\hat{y}^*)^T \end{bmatrix} = J_{f'}^T(a, x, y)\lambda^T \quad (32)$$

lead to solutions of equations (27)–(30) by homotopy paths. Each homotopy path is the set of zeros of the straight-line homotopy function of $t \in (0, 1]$ as t varies from 1 to 0:

$$H(x, t) = \begin{cases} f'(a, x, y) \\ \begin{bmatrix} 0 \\ y^T - (t(\hat{y}^*)^T + (1-t)\hat{y}^T) \end{bmatrix} - J_{f'}^T(a, x, y)\lambda^T \end{cases}.$$

As $t \rightarrow 0$, we obtain numerical approximations to the solutions of equations (27)–(30).

The benefit of employing a parameter homotopy is that after solving equations (31)–(32) with a more general method *a priori*, we significantly reduce the computation required in solving equations (27)–(30) for each i . In practice, the entries of \hat{y}^* are sampled uniformly along the complex unit circle. More details on how parameter homotopies fit into numerical algebraic geometry can be found in [1] and [2].

In addition to employing a parameter homotopy solving scheme, equations (31)–(32), or equally equations (27)–(30), have a natural $\{(a, x, y), \lambda\}$ -homogenous structure (see [1] and [2]). This observation significantly reduces the number of homotopy paths that need to be tracked numerically. In addition, this increases stability of path tracking. Multihomogenous structures are used alongside parameter homotopies to solve equations (27)–(30).

One additional reformulation that we can do to reduce computation is to define some of the variables intrinsically. This is common if one or more variables can be written as a linear combination of some of the other variables. Specifically, we know from Table 7 that:

$$x_2 = \text{cRAF}_{tot}$$

where cRAF_{tot} is defined as a constant in Table 8. Thus, we can “remove” x_2 from our computations. Partial derivatives are no longer necessary with respect to the variable x_2 , and x_2 is no longer defined explicitly when tracking homotopy paths.

Table 9 summarizes the sequence of reductions made in the number of paths by imposing a $\{(a, x, y), \lambda\}$ -homogeneous structure followed by intrinsically defining the variable x_2 along with the number of paths required using the standard total degree homotopy [1],[2].

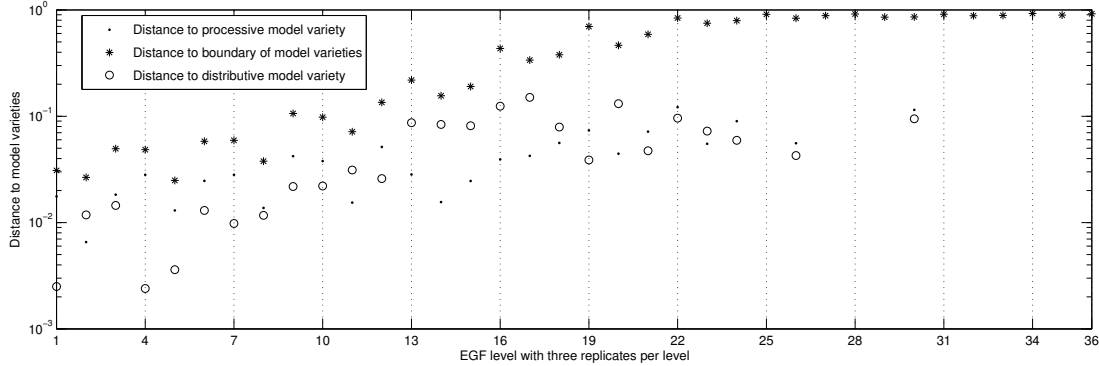
Table 10: Expected timings for the MAPK model collected over 20 ‘random’ runs.

	Compute Dimension	Initial Solve (parallel)	Data Solve (all 36)
Processive Model	9.14 sec \pm 1.21 sec	28.05 sec \pm 3.24 sec	10.95 sec \pm 2.09 sec
Distributive Model	13.54 sec \pm 1.08 sec	53.06 sec \pm 9.20 sec	19.09 sec \pm 4.84 sec

Table 12 and Table 13 record the distances between the data and model varieties for all 36 data points. A missing “interior” distance in Table 12 and Table 13 indicate there were no positive real critical points found for the given EGF level and replicate. However, we may still compute a distance to the boundary of the semi-algebraic sets corresponding to each model. **Bertini** input files, shell scripts, and **MATLAB** scripts are available within the supplementary files to analyze model selection and parameter estimation. The distances are summarized graphically in Figure S1.

Timing summaries for both the processive and distributive model can be found in Table 10. These timings include the numerical irreducible decomposition required to compute the dimension of each component in the model variety, solving equations (31)–(32) required to employ a parameter homotopy scheme, and the parameter homotopy to solve equations (27)–(30) for all i . Timings to compute the dimension of the model variety and the data solve were done in serial using a Apple MacBook Pro with 2.4 GHz Intel “Core i5” processor. The initial solve for the parameter homotopies were done in parallel using 96 (2.67 GHz Xeon-5650) compute nodes on the CentOS 5.11 operating system.

Figure S1: Distance Plot



References

- [1] Sommese AJ, Wampler CW (2005) *The numerical solution to systems of polynomials arising in engineering and science* (World Scientific, Singapore).
- [2] Bates DJ, Hauenstein JD, Sommese AJ, Wampler CW (2013) *Numerically solving polynomial systems with Bertini* (SIAM) Vol. 25.
- [3] Wu W, Reid G (2013) Finding points on real solution components and applications to differential polynomial systems. *ISSAC '13 Proceedings of the 38th International Symposium on Symbolic and Algebraic Computation* 16:339–346.

- [4] Rouillier F, Roy MF, Din MSE (2000) Finding at least one point in each connected component of a real algebraic set defined by a single equation. *Journal of Complexity* 16:716–750.
- [5] Hauenstein JD (2013) Numerically computing real points on algebraic sets. *Acta Applicandae Mathematicae* 125:105–119.
- [6] Verschelde J (1999) Algorithm 795: Phcpack: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Transactions on Mathematical Software (TOMS)* 25:251–276.
- [7] Lee TL, Li TY, Tsai CH (2008) Hom4ps-2.0: a software package for solving polynomial systems by the polyhedral homotopy continuation method. *Computing* 83:109–133.
- [8] Bates DJ, Hauenstein JD, Sommese AJ, Wampler CW (2014) Bertini: Software for numerical algebraic geometry. available at bertini.nd.edu.
- [9] Rostalski P, Fotiou I, Bates D, Beccuti A, Morari M (2011) Numerical algebraic geometry for optimal control applications. *SIAM Journal on Control and Optimization* 21:417–437.
- [10] Gross E, Harrington HA, Rosen Z, Sturmfels B (2016) Algebraic systems biology: A case study for the wnt pathway. *Bulletin of Mathematical Biology* 78:21–51.
- [11] Scott FL, et al. (2009 Feb 19) The fas-fadd death domain complex structure unravels signalling by receptor clustering. *Nature* 457:1019–1022.
- [12] Siegal-Gaskins D, Franco E, Zhou T, Murray RM (2014) An analytical approach to bistable biological circuit discrimination using real algebraic geometry. *BioRxiv*.
- [13] Gardner T, Cantor C, JJ C (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature* 403:339–342.
- [14] Hasty J, Isaacs F, Dolnik M, McMillen D, Collina J (2001) Designer gene networks: Towards fundamental cellular control. *Chaos* 11:207–220.
- [15] Siegal-Gaskins D, Mejia-Guerra M, Smith G, E G (2011) Emergence of switch-like behavior in a large family of simple biochemical networks. *PLoS Computational Biology* 7:e1002039.
- [16] Seidenberg A (1954) A new decision method for elementary algebra. *Ann. of Math.* 2:365–374.
- [17] Hernandez-Vargas EA, Middleton R, Mehta D (2011) Towards modeling hiv long term behavior. *Proceedings of the 18th IFAC World Congress* 18:581–586.
- [18] Hadjiandreou M, Conejeros R, Vassiliadis V (2007) Towards a long-term model construction for the dynamic simulation of hiv infection. *Mathematical Bioscience and Engineering* 4:489–504.
- [19] Aoki K, Yamada M, Kunida K, Yasuda S, Matsuda M (2011) Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* 108:12675–12680.
- [20] D. Bates, E. Gross AL, Rodriguez J (year?) Bertini for macaulay2 available at [arxiv:1310.3297](https://arxiv.org/abs/1310.3297).
- [21] Grayson D, Stillman M (year?) *Macaulay2*, a software system for research in algebraic geometry, available at <http://www.math.uiuc.edu/macaulay2/>.

Table 11: Parameter estimate of k_1 for distributive and processive MAP Kinase models

EGF level	Distributive	Processive
1 (0.0244140625 ng/mL)	0.006655185015566	0.002630893498837
1	0.005169208080985	0.002666996926268
1	0.010517845922915	0.004869916688582
2 (0.048828125 ng/mL)	0.010599752816972	0.004139244229281
2	0.005294859090859	0.002185712163548
2	0.012645415710605	0.005598240936340
3 (0.09765625 ng/mL)	0.013040547423470	0.005555450037129
3	0.007862190037618	0.003676690633723
3	0.007862190037618	0.010890090940375
4 (0.1953125 ng/mL)	0.022314241866226	0.007566455646161
4	0.014767039426564	0.010026925643431
4	0.032112677267837	0.014358973830327
5 (0.390625 ng/mL)	0.057037089355901	0.028188983627261
5	0.034598433900385	0.018615955020805
5	0.046993978170041	0.023610675947125
6 (0.78125 ng/mL)	0.171132616846834	0.081810937526556
6	0.108600436914432	0.052541291660911
6	0.128469450822607	0.062127115025643
7 (1.5625 ng/ML)	0.552602449693322	0.311829951745710
7	0.198177806441869	0.094130793284512
7	0.307630980846653	0.162410456627761
8 (3.125 ng/ML)	1.535918937663066	1.104298831092584
8	1.558792683503375	0.653311235583287
8	1.114642498639051	0.700052847271085
9 (6.25 ng/ML)	0	0
9	6.741089632275663	2.682148283074403
9	0	0
10 (12.5 ng/ML)	0	0
10	0	0
10	2.556601780467115	1.856045810178439
11 (25 ng/ML)	0	0
11	0	0
11	0	0
11	0	0
12 (50 ng/ML)	0	0
12	0	0
12	0	0

Table 12: Distance to (smaller) distributive model variety

EGF level	Replicate	“Interior” distance	“Boundary” distance
1	1	0.0025	0.0309
	2	0.0118	0.0266
	3	0.0145	0.0496
2	1	0.0024	0.0485
	2	0.0036	0.0249
	3	0.0130	0.0581
4	1	0.0098	0.0594
	2	0.0117	0.0377
	3	0.0218	0.1062
8	1	0.0221	0.0981
	2	0.0312	0.0714
	3	0.0259	0.1349
16	1	0.0870	0.2189
	2	0.0838	0.1559
	3	0.0814	0.1904
32	1	0.1243	0.4343
	2	0.1505	0.3374
	3	0.0791	0.3784
64	1	0.0388	0.6990
	2	0.1312	0.4648
	3	0.0473	0.5889
128	1	0.0959	0.8398
	2	0.0725	0.7501
	3	0.0594	0.7931
256	1	—	0.9093
	2	0.0427	0.8353
	3	—	0.8839
512	1	—	0.9154
	2	—	0.8556
	3	0.0947	0.8597
1024	1	—	0.9111
	2	—	0.8817
	3	—	0.8883
2048	1	—	0.9272
	2	—	0.8948
	3	—	0.9197

Table 13: Distance to (larger) processive model variety

EGF level	Replicate	“Interior” distance	“Boundary” distance
1	1	0.0176	0.0309
	2	0.0066	0.0266
	3	0.0183	0.0496
2	1	0.0281	0.0485
	2	0.0130	0.0249
	3	0.0247	0.0581
4	1	0.0282	0.0594
	2	0.0137	0.0377
	3	0.0421	0.1062
8	1	0.0379	0.0981
	2	0.0154	0.0714
	3	0.0514	0.1349
16	1	0.0284	0.2189
	2	0.0156	0.1559
	3	0.0246	0.1904
32	1	0.0392	0.4343
	2	0.0424	0.3374
	3	0.0561	0.3784
64	1	0.0735	0.6990
	2	0.0444	0.4648
	3	0.0717	0.5889
128	1	0.1218	0.8398
	2	0.0550	0.7501
	3	0.0899	0.7931
256	1	—	0.9093
	2	0.0557	0.8353
	3	—	0.8839
512	1	—	0.9154
	2	—	0.8556
	3	0.1149	0.8597
1024	1	—	0.9111
	2	—	0.8817
	3	—	0.8883
2048	1	—	0.9272
	2	—	0.8948
	3	—	0.9197